BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

4/89 -

PCT/AU97/00124

5

REC'D 2 4 MAR 1997	
WIPO	PCT

Patent Office
Canberra

I, DAVID DANIEL CLARKE , ASSISTANT DIRECTOR PATENT SERVICES, hereby certify that the annexed are true copies of the provisional specification and drawing(s) as filed on 1 March 1996 in connection with Application No. PN 8386 for a patent by FLORIGENE PTY LTD filed on 1 March 1996.

I further certify that the name of the applicant has been amended to FLORIGENE LIMITED pursuant to the provisions of Section 104 of the Patents Act 1990.

I further certify that the annexed documents are not, as yet, open to public inspection.

PRIORITY DOCUMENT

WITNESS my hand this Thirteenth
day of March 1997

DAVID DANIEL CLARKE
ASSISTANT DIRECTOR PATENT SERVICES

Regulation 3.2

FLORIGENE LIMITED

AUSTRALIA
Patents Act 1990

PROVISIONAL SPECIFICATION
for the invention entitled:

"Genetic sequences encoding flavonoid pathway enzymes and uses therefor"

The invention is described in the following statement:

GENETIC SEQUENCES ENCODING FLAVONOID PATHWAY ENZYMES AND USES THEREFOR

The present invention relates generally to genetic sequences encoding flavonoid pathway metabolising enzymes and more particularly to flavonoid 3'-hydroxylase or derivatives thereof and their use in the manipulation of pigmentation in plants and other organisms.

Bibliographic details of the publications referred to by the author in this specification are collected at the end of the description. Sequence Identity Numbers (SEQ ID NOs.) for the nucleotide and amino acid sequences referred to in the specification are defined following the bibliography.

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element or integer or group of elements or integers but not the exclusion of any other element or integer or group of elements or integers.

The rapidly developing sophistication of recombinant DNA technology is greatly facilitating research and development in a range of biotechnology related industries. The horticultural industry has become a recent beneficiary of this technology which has contributed to developments in disease resistance in plants and flowers exhibiting delayed senescence after cutting. Some attention has also been directed to manipulating flower colour.

The flower industry strives to develop new and different varieties of flowering plants. An effective way to create such novel varieties is through the manipulation of flower colour. Classical breeding techniques have been used with some success to produce a wide range of colours for most of the commercial varieties of flowers. This approach has been limited, however, by the constraints of a particular species' gene pool and for this reason it is rare for a single species to have a full spectrum of coloured varieties.

In addition, traditional breeding techniques lack precision. The aesthetic appeal of the flower is a combination of many factors such as form, scent and colour; modification of one character through hybridization can often be at the expense of an equally valuable feature. The ability to genetically engineer precise colour changes in cutflower and ornamental species would offer significant commercial opportunities in an industry which has rapid product turnover and where novelty is an important market characteristic.

Flower colour is predominantly due to two types of pigment: flavonoids and carotenoids. Flavonoids contribute to a range of colours from yellow to red to blue. Carotenoids impart an orange or yellow tinge and are commonly the major pigment in yellow or orange flowers. The flavonoid molecules which make the major contribution to flower colour are the anthocyanins which are glycosylated derivatives of cyanidin, delphinidin, petunidin, peonidin, malvidin and pelargonidin, and are localised in the vacuole. The different anthocyanins can produce marked differences in colour. Flower colour is also influenced by co-pigmentation with colourless flavonoids, metal complexation, glycosylation, acylation and vacuolar pH (Forkmann, 1991).

The biosynthetic pathway for the flavonoid pigments (hereinafter referred to as the "flavonoid pathway") is well established and is shown in Figures 1a and 1b (Ebel and Hahlbrock, 1988; Hahlbrock and Grisebach, 1979; Wiering and De Vlaming, 1984; Schram *et al.*, 1984; Stafford, 1990; Van Tunen and Mol, 1990; Dooner *et al*, 1991; Martin and Gerats, 1993; Holton and Cornish, 1995). The first committed step in the pathway involves the condensation of three molecules of malonyl-CoA with one molecule of *p*-coumaroyl-CoA. This reaction is catalysed by the enzyme chalcone synthase (CHS). The product of this reaction, 2',4,4',6', tetrahydroxy-chalcone, is normally rapidly isomerized to produce naringenin by the enzyme chalcone flavanone isomerase (CHI). Naringenin is subsequently hydroxylated at the 3 position of the central ring by flavanone 3-hydroxylase (F3H) to produce dihydrokaempferol (DHK).

The pattern of hydroxylation of the B-ring of DHK plays a key role in determining petal colour. The B-ring can be hydroxylated at either the 3′, or both the 3′ and 5′ positions, to produce dihydroquercetin (DHQ) and dihydromyricetin (DHM), respectively. Two key enzymes involved in this pathway are flavonoid 3′-hydroxylase and flavonoid 3′,5′-hydroxylase, both of the cytochrome P450 class. Cytochrome P450 enzymes are widespread in nature and genes have been isolated and sequenced from vertebrates, insects, yeasts, fungi, bacteria and plants.

Flavonoid 3′-hydroxylase acts on DHK to produce DHQ and on naringenin to produce eriodictyol. Reduction and glycosylation of DHQ produces the cyanidin-glycoside and peonidin-glycoside pigments which, in many plant species (for example rose, carnation and chrysanthemum), contribute to red and pink flower colour. The synthesis of these anthocyanins can also result in other flower colours. For example, blue cornflowers contain cyanin. The ability to control flavonoid 3′-hydroxylase activity, or other enzymes involved in the flavonoid pathway, in flowering plants would provide a means to manipulate petal colour. Different coloured versions of a single cultivar could thereby be generated and in some instances a single species would be able to produce a broader spectrum of colours.

A genetic sequence encoding a petunia flavonoid 3′-hydroxylase has been cloned (see International Patent Application No. PCT/AU93/00127 [WO 93/20206]). However, this sequence was inefficient in modulation of 3′-hydroxylated anthocyanins in plants. There is a need, therefore, to develop further procedures for cloning genetic sequences encoding flavonoid 3′-hydroxylases and to clone these sequences-from plants.

In accordance with the present invention, genetic sequences encoding flavonoid 3′-hydroxylase have been identified and cloned. The recombinant genetic sequences of the present invention permit the modulation of expression of genes encoding this enzyme by, for example, *de novo* expression, over-expression, suppression, antisense inhibition and ribozyme activity. The ability to control flavonoid 3′-hydroxylase synthesis permits modulation of the composition of individual anthocyanins as well as

alteration of relative levels of flavonols and anthocyanins, thereby enabling the manipulation of petal colour.

Accordingly, one aspect of the present invention provides a nucleic acid molecule comprising a nucleotide sequence or complementary nucleotide sequence substantially as set forth in SEQ ID NO:1 or having at least 60% similarity thereto or capable of hybridising to the sequence set forth in SEQ ID NO:1 under low stringency conditions.

In a related embodiment, there is provided a nucleic acid molecule comprising a nucleotide sequence or complementary nucleotide sequence substantially as set forth in SEQ ID NO:3 or having at least 60% similarity thereto or capable of hybridising to the sequence set forth in SEQ ID NO:3 under low stringency conditions.

In another related embodiment, the present invention is directed to a nucleic acid molecule comprising a nucleotide sequence or complementary nucleotide sequence substantially as set forth in SEQ ID NO:5 or having at least 60% similarity thereto or capable of hybridising to the sequence set forth in SEQ ID NO:5 under low stringency conditions.

Yet another related embodiment provides a nucleic acid molecule comprising a nucleotide sequence or complementary nucleotide sequence substantially as set forth in SEQ ID NO:7 or having at least 60% similarity thereto or capable of hybridising to the sequence set forth in SEQ ID NO:7 under low stringency conditions.

Still yet a further embodiment of the present invention relates to a nucleic acid molecule comprising a nucleotide sequence or complementary nucleotide sequence substantially as set forth in SEQ ID NO:9 or having at least 60% similarity thereto or capable of hybridising to the sequence set forth in SEQ ID NO:9 under low stringency conditions.

Preferably, in relation to the aspects of the present invention, the percentage similarity is at least about 70%, more preferably at least about 80% and still more preferably at least about 90-95%.

Reference herein to a low stringency at 42°C includes and encompasses from at least about 1% to at least about 15% formamide and from at least about 1M to at least about 2M salt for hybridization, and at least about 1M to at least about 2M salt for washing conditions. Alternative stringency conditions may be applied where necessary, such as medium stringency, which includes and encompasses from at least about 16% to at least about 30% formamide and from at least about 0.5M to at least about 0.9M salt for hybridization, and at least about 0.5M to at least about 0.9M salt for washing conditions, or high stringency, which includes and encompasses from at least about 31% to at least about 50% formamide and from at least about 0.01M to at least about 0.15M salt for hybridization, and at least about 0.01M to at least about 0.15M salt for washing conditions.

Another embodiment of the present invention provides a nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence substantially as set forth in SEQ ID NO:2 or an amino acid sequence having at least about 50% similarity thereto.

In a related embodiment, there is provided a nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence substantially as set forth in SEQ ID NO:4 or an amino acid sequence having at least about 50% similarity thereto.

A further related embodiment of the present invention is directed to a nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence substantially as set forth in SEQ ID NO:6 or an amino acid sequence having at least about 50% similarity thereto.

Still another related embodiment provides a nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence substantially as set forth in SEQ ID NO:8 or an amino acid sequence having at least about 50% similarity thereto.

Yet still another related embodiment relates to a nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence substantially as set forth in SEQ ID NO:10 or an amino acid sequence having at least about 50% similarity thereto.

Preferably, in accordance with the embodiments of the present invention the percentage similarity is at about 60%, more preferably at least about 70%, even more preferably at least about 80% and still more preferably at least about 90-95%.

The nucleic acid molecule defined by SEQ ID NO:1 encodes a flavonoid 3'-hydroxylase (F3'H) from petunia. Examples of other suitable F3'H genes are from snapdragon (SEQ ID NO:3), arabidopsis (SEQ ID NO:5), carnation (SEQ ID NO:7) and rose (SEQ ID NO:9). Although the present invention is particularly exemplified by the aforementioned F3'H genes, the subject invention extends to F3'H genes from any species of plant provided that the F3'H gene has at least about 60% similarity at the nucleotide level and/or at least about 50% similarity at the amino acid level to a nucleic acid molecule selected from SEQ ID NO:1 or 3 or 5 or 7 or 9 or SEQ ID NO: 2 or 4 or 6 or 8 or 10, respectively.

The nucleic acid molecules of the present invention are generally genetic sequences in a non-naturally-occurring condition. Generally, this means isolated away from its natural state or synthesized or derived in a non-naturally-occurring environment. More specifically, it includes nucleic acid molecules formed or maintained *in vitro,* including genomic DNA fragments, recombinant or synthetic molecules and nucleic acids in combination with heterologous nucleic acids. It also extends to the genomic DNA or cDNA or part thereof encoding F3'H or part thereof in reverse orientation relative to its or another promoter. It further extends to naturally-occurring sequences following at least a partial purification relative to other nucleic acid sequences.

The term "nucleic acid molecule" includes a nucleic acid isolate and a genetic sequence and is used herein in its most general sense and encompasses any contiguous series of nucleotide bases specifying directly, or *via* a complementary series of bases, a sequence

of amino acids in a F3'H. Such a sequence of amino acids may constitute a full-length F3'H or an active truncated form thereof or may correspond to a particular region such as an N-terminal, C-terminal or internal portion of the enzyme. The nucleic acid molecules contemplated herein also encompass oligonucleotides useful as genetic probes or as "antisense" molecules capable of regulating expression of the corresponding gene in a plant. An "antisense molecule" as used herein may also encompass a gene construct comprising the structural genomic or cDNA gene or part thereof in reverse orientation relative to its or another promoter. Accordingly, the nucleic acid molecules of the present invention may be suitable for use as cosuppression molecules, ribozyme molecules, sense molecules and antisense molecules to modulate levels of 3'-hydroxylated anthocyanins.

In one embodiment, the nucleic acid molecule encoding F3'H or various derivatives thereof are used to reduce the activity of an endogenous F3'H, or alternatively the nucleic acid molecule encoding this enzyme or various derivatives or parts thereof are used in the antisense orientation to reduce activity of the F3'H. Although not wishing to limit the present invention to any one theory, it is possible that the introduction of the nucleic acid molecule into a cell results in this outcome either by decreasing transcription of the homologous endogenous gene or by increasing turnover of the corresponding mRNA. This may be achieved using gene constructs containing F3'H nucleic acid molecules or various derivatives or parts thereof in either the sense or the antisence orientation. In a further alternative, ribozymes could be used to inactivate target nucleic acid molecules. Alternatively, the nucleic acid molecule encodes a functional F3'H and this is used to elevate levels of this enzyme.

Reference herein to the altering of flavonoid F3'H activity relates to an elevation or reduction in activity of up to 30% or more preferably of 30-50%, or even more preferably 50-75% or still more preferably 75% or greater above or below the normal endogenous or existing levels of activity. The level of activity can be readily assayed using a modified version of the method described by Stotz and Forkmann (1982) (see Example 7) or by assaying for the amount of F3'H product such as quercetin, cyanidin or peonidin as set forth in Example 5.

The present invention further extends to nucleic acid molecules in the form of oligonucleotide primers or probes capable of hybridizing to a portion of the nucleic acid molecules contemplated above, and in particular those selected from the nucleic acid molecules set forth in SEQ ID NOs: 1, 3, 5, 7 and 9, under high, preferably under medium and most preferably under low stringency conditions. Preferably the portion corresponds to the 5′ or the 3′ end of the F3′H gene. For convenience the 5′ end is considered herein to define a region substantially between the 5′ end of the primary transcript to a centre portion of the gene, and the 3′ end is considered herein to define a region substantially between the centre portion of the gene and the 3′ end of the primary transcript. It is clear, therefore, that oligonucleotides or probes may hybridize to the 5′ end or the 3′ end or to a region common to both the 5′ and the 3′ ends.

The nucleic acid molecule or its complementary form may encode the full-length enzyme or a part or derivative thereof. By "derivative" is meant any single or multiple amino acid substitutions, deletions, and/or additions relative to the naturally-occurring enzyme and includes parts, fragments, portions, homologues and analogues. In this regard, the nucleic acid includes the naturally-occurring nucleotide sequence encoding F3′H or may contain single or multiple nucleotide substitutions, deletions and/or additions to said naturally-occurring sequence. The nucleic acid of the present invention or its complementary form may also encode a "part" of the F3′H, whether active or inactive, and such a nucleic acid molecule may be useful as an oligonucleotide probe, primer for polymerase chain reactions or in various mutagenic techniques, for the generation of antisense molecules or in the construction of ribozymes. They may also be useful in developing co-suppression constructs. The nucleic acid molecule of the present invention may or may not encode a functional F3′H.

Amino acid insertional derivatives of the F3′H of the present invention include amino and/or carboxyl terminal fusions as well as intra-sequence insertions of single or multiple amino acids. Insertional amino acid sequence variants are those in which one or more amino acid residues are introduced into a predetermined site in the protein although random insertion is also possible with suitable screening of the resulting product. Deletional variants are characterised by the removal of one or more amino

acids from the sequence. Substitutional amino acid variants are those in which at least one residue in the sequence has been removed and a different residue inserted in its place. Typical substitutions are those made in accordance with Table 1 below.

TABLE 1

Suitable residues for amino acid substitutions

Original Residue	Exemplary Substitutions
Ala	Ser
Arg	Lys
Asn	Gln; His
Asp	Glu
Cys	Ser
Gln	Asn
Glu	Asp
Gly	Pro
His	Asn; Gln
Ile	Leu; Val
Leu	Ile; Val
Lys	Arg; Gln; Glu
Met	Leu; Ile
Phe	Met; Leu; Tyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp; Phe
Val	Ile; Leu

Where the F3'H is derivatised by amino acid substitution, the amino acids are generally replaced by other amino acids having like properties, such as hydrophobicity, hydrophilicity, electronegativity, bulky side chains and the like. Amino acid substitutions are typically of single residues. Amino acid insertions will usually be in

the order of about 1-10 amino acid residues and deletions will range from about 1-20 residues. Preferably, deletions or insertions are made in adjacent pairs, i.e. a deletion of two residues or insertion of two residues.

The amino acid variants referred to above may readily be made using peptide synthetic techniques well known in the art, such as solid phase peptide synthesis (Merrifield, 1964) and the like, or by recombinant DNA manipulations. Techniques for making substitution mutations at predetermined sites in DNA having known or partially known sequence are well known and include, for example, M13 mutagenesis. The manipulation of DNA sequence to produce variant proteins which manifest as substitutional, insertional or deletional variants are conveniently described, for example, in Sambrook *et al.* (1989).

Other examples of recombinant or synthetic mutants and derivatives of the F3'H of the present invention include single or multiple substitutions, deletions and/or additions of any molecule associated with the enzyme such as carbohydrates, lipids and/or proteins or polypeptides.

The terms "analogues" and "derivatives" also extend to any functional chemical equivalent of the F3'H and also to any amino acid derivative described above. For convenience, reference to "F3'H" herein includes reference to any mutants, derivatives, analogues, homologues or fragments thereof.

The present invention is exemplified using nucleic acid sequences derived from petunia, carnation, rose, snapdragon and arabidopsis since these represent the most convenient and preferred source of material to date. However, one skilled in the art will immediately appreciate that similar sequences can be isolated from any number of sources such as other plants or certain microorganisms. Examples of other plants include, but are not limited to, chrysanthemum, maize, tobacco, cornflower, pelargonium, morning glory, apple, gerbera and african violet. All such nucleic acid sequences encoding directly or indirectly a flavonoid pathway enzyme and in particular F3'H, regardless of their source, are encompassed by the present invention.

The nucleic acid molecules contemplated herein may exist in either orientation alone or in combination with a vector molecule, for example an expression-vector. The term vector molecule is used in its broadest sense to include any intermediate vehicle for the nucleic acid molecule, capable of facilitating transfer of the nucleic acid into the plant cell and/or facilitating integration into the plant genome. An intermediate vehicle may, for example, be adapted for use in electroporation, microprojectile bombardment, *Agrobacterium*-mediated transfer or insertion via DNA or RNA viruses. The intermediate vehicle and/or the nucleic acid molecule contained therein may or may not need to be stably integrated into the plant genome. Such vector molecules may also replicate and/or express in prokaryotic cells. Preferably, the vector molecules or parts thereof are capable of integration into the plant genome. The nucleic acid molecule may additionally contain a promoter sequence capable of directing expression of the nucleic acid molecule in a plant cell. The nucleic acid molecule and promoter may also be introduced into the cell by any number of means such as those described above.

In accordance with the present invention, a nucleic acid molecule encoding a F3'H or a derivative or part thereof may be introduced into a plant in either orientation to allow, permit or otherwise facilitate manipulation of levels of production of mRNA in the cytoplasm and/or production of enzyme from the mRNA, thereby providing a means either to convert DHK and/or other suitable substrates, if synthesised in the plant cell, ultimately into anthocyanin derivatives of anthocyanidins such as cyanidin and/or peonidin, or alternatively to inhibit such conversion of metabolites by reducing or eliminating endogenous or existing F3'H activity. The manipulation of levels of production of mRNA in the cytoplasm and/or production of enzyme from the mRNA, is referred to herein as "expression". The production of anthocyanins contributes to the production of a red or blue flower colour. Expression of the nucleic acid molecule in either orientation in the plant may be constitutive, inducible or developmental, and may also be tissue-specific.

According to this aspect of the present invention there is provided a method for producing a transgenic plant capable of synthesizing F3'H or functional derivatives thereof, said method comprising stably transforming a cell of a suitable plant with a

nucleic acid molecule which comprises a sequence of nucleotides encoding said F3'H, under conditions permitting the eventual expression of said nucleic acid molecule, regenerating a transgenic plant from the cell and growing said transgenic plant for a time and under conditions sufficient to permit the expression of the nucleic acid molecule. The transgenic plant may thereby produce elevated levels of F3'H activity relative to the amount expressed in a comparable non-transgenic plant.

Another aspect of the present invention contemplates a method for producing a transgenic plant with reduced endogenous or existing F3'H activity, said method comprising stably transforming a cell of a suitable plant with a nucleic acid molecule which comprises a sequence of nucleotides encoding or complementary to a sequence encoding F3'H, regenerating a transgenic plant from the cell and where necessary growing said transgenic plant under conditions sufficient to permit the expression of the nucleic acid molecule.

Yet another aspect of the present invention contemplates a method for producing a genetically modified plant with reduced endogenous or existing F3'H activity, said method comprising altering the F3'H gene through modification of the endogenous sequences via homologous recombination from an appropriately altered F3'H gene or derivative or part thereof introduced into the plant cell, and regenerating the genetically modified plant from the cell.

In accordance with these aspects of the present invention the preferred nucleic acid molecules are substantially as set forth in SEQ ID NO:1, 3, 5, 7 or 9 or have at least about 60% similarity thereto or are capable of hybridising thereto under low stringency conditions.

In a preferred embodiment, the present invention contemplates a method for producing a transgenic flowering plant exhibiting altered flower colour, said method comprising stably transforming a cell of a suitable plant with a nucleic acid molecule of the present invention, regenerating a transgenic plant from the cell and growing said transgenic plant for a time and under conditions sufficient to permit the expression of the nucleic

acid molecule into the F3'H enzyme. Alternatively, said method may comprise stably transforming a cell of a suitable plant with a nucleic acid molecule of the present invention or its complementary sequence, regenerating a transgenic plant from the cell and growing said transgenic plant for a time and under conditions sufficient to alter the level of activity of the endogenous or existing F3'H. Preferably, the altered level would be less than the endogenous or existing level of F3'H activity in a comparable non-transgenic plant.

In a related embodiment, the present invention contemplates a method for producing a flowering plant exhibiting altered flower colour, said method comprising alteration of the F3'H gene through modification of the endogenous sequences *via* homologous recombination from an appropriately altered F3'H gene or derivative thereof introduced into the plant cell and regenerating the genetically modified plant from the cell.

The nucleic acid molecules of the present invention may or may not be developmentally regulated. Preferably, the modulation of levels of 3'-hydroxylated anthocyanins leads to altered flower colour which includes the production of red flowers or other colour shades depending on the physiological conditions of the recipient plant. By "recipient plant" is meant a plant capable of producing a substrate for the F3'H enzyme, or producing the F3'H enzyme itself, and possessing the appropriate physiological properties and genotype required for the development of the colour desired. This may include but is not limited to petunia, carnation, chrysanthemum, rose, snapdragon, tobacco, cornflower, pelargonium, lisianthus, gerbera, apple, iris, lily, african violet and morning glory.

Accordingly, the present invention extends to a method for producing a transgenic plant capable of modulating levels of 3'-hydroxylated anthocyanins, said method comprising stably transforming a cell or group of cells of a suitable plant with a nucleic acid molecule comprising a sequence of nucleotides encoding, or complementary to a sequence encoding, F3'H or a derivative thereof, and regenerating a transgenic plant from said cell or cells.

One skilled in the art will immediately recognise the variations applicable to the methods of the present invention, such as increasing or decreasing the level of enzyme activity of the enzyme naturally present in a target plant leading to differing shades of colours.

The present invention, therefore, extends to all transgenic plants containing all or part of the nucleic acid module of the present invention and/or any homologues or related forms thereof or antisense forms of any of these and in particular those transgenic plants which exhibit altered flower colour. The transgenic plants may contain an introduced nucleic acid molecule comprising a nucleotide sequence encoding or complementary to a sequence encoding F3'H. Generally, the nucleic acid would be stably introduced into the plant genome, although the present invention also extends to the introduction of the F3'H nucleotide sequence within an autonomously-replicating nucleic acid sequence such as a DNA or RNA virus capable of replicating within the plant cell. The invention also extends to seeds from such transgenic plants. Such seeds, especially if coloured, will be useful as proprietary tags for plants.

A further aspect of the present invention is directed to recombinant forms of F3'H. The recombinant forms of the enzymes will provide a source of material for research to develop, for example, more active enzymes and may be useful in developing *in vitro* systems for production of coloured compounds.

Still a further aspect of the present invention contemplates the use of the genetic sequences described herein in the manufacture of a genetic construct capable of use in modulating levels of 3'-hydroxylated anthocyanins in a plant or cells of a plant.

Yet a further aspect of the present invention provides flowers and in particular cut flowers, from the transgenic plants herein described, exhibiting altered flower colour.

Another aspect of the present invention is directed to a nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding, a F3'H or a derivative thereof wherein said nucleic acid molecule is capable

of being expressed in a plant cell. The term "expressed" is equivalent to the term "expression" as defined above.

The nucleic acid molecules according to this and other aspects of the invention allow, permit or otherwise facilitate increased efficiency in modulation of 3′-hydroxylated anthocyanins relative to the efficency of the pCGP619 cDNA insert contained in plasmid pCGP809, disclosed in International Patent Application No. PCT/AU93/00127 [WO 93/20206]. The term "plant cell" includes a single plant cell or a group of plant cells such as in a callus, plantlet or plant or parts thereof including flowers and seeds.

The present invention is further described by reference to the following non-limiting Figures and Examples.

In the Figures:

Figures 1a and 1b Schematic representation of the flavonoid biosynthesis pathways in *P. hybrida* flowers showing the enzymes and genetic loci involved in the conversions. Enzymes involved in the pathway have been indicated as follows: PAL = phenylalanine ammonia-lyase; C4H = cinnamate 4-hydroxylase; 4CL = 4-coumarate: CoA ligase; CHS = chalcone synthase; CHI = chalcone isomerase; F3H = flavanone 3-hydroxylase; F3′H = flavonoid 3′-hydroxylase; F3′5′H = flavonoid 3′5′ hydroxylase; FLS = flavonol synthase; DFR = dihydroflavonol-4-reductase; ANS = anthocyanin synthase; 3GT = UDP-glucose: anthocyanin-3-glucoside; 3RT = UDP-rhamnose: anthocyanidin-3-glucoside rhamnosyltransferase; ACT = anthocyanidin-3-rutinoside acyltransferase; 5GT = UDP-glucose: anthocyanin 5- glucosyltransferase; 3′ OMT = anthocyanin *O*-methyltransferase; 3′, 5′ OMT = anthocyanin 3′, 5′ *O*-methyltransferase. The flavonol, myricetin is only produced at low levels and the anthocyanin, pelargonidin is rarely produced in *P. hybrida*.

Figure 2 Diagrammatic representation of the plasmid pCGP161 containing a cDNA clone representing the cinnamate-4-hydroxylase from *P. hybrida*. ^{32}P-labelled fragments of the 0.7 kb *Eco*RI/*Xho*I fragment were used to probe the Old Glory Red

petal cDNA library.

Figure 3 Diagrammatic representation of the plasmid pCGP602 containing a cDNA clone (617) representing a flavonoid 3'5' hydroxylase (*Hf1*) from *P. hybrida*. ^{32}P-labelled fragments of the 1.6 kb *BspHI*/*Fsp*I fragment containing the *Hf1* coding region were used to probe the Old Glory Red petal cDNA library.

Figure 4 Diagrammatic representation of the plasmid pCGP175 containing a cDNA clone (H2) representing a flavonoid 3'5' hydroxylase (*Hf2*) from *P. hybrida*. ^{32}P-labelled fragments of the 1.3 kb *EcoR*I/*Xho*I and 0.5 kb *Xho*I fragments which together contain the *Hf2* coding region were used to probe the Old Glory Red petal cDNA library.

Figure 5 Diagrammatic representation of the plasmid pCGP619 containing the 651 cDNA clone representing a cytochrome P450 from *P. hybrida*. ^{32}P-labelled fragments of the 1.8 kb *EcoR*I/*Xho*I fragment were used to probe the Old Glory Red petal cDNA library.

Figure 6 Representation of an autoradiograph of an RNA blot probed with ^{32}P-labelled fragments of the OGR-38 cDNA clone contained in pCGP1805 (Figure 8). Each lane contained a 20 μg sample of total RNA isolated from the flowers or leaves of plants of a V23 (ht1/ht1) x VR (Ht1/ht1) backcross population. A 1.8 kb transcript was detected in the VR-like (Ht1/ht1) flowers that contained high levels of quercetin (Q+)(lanes 9 - 14). The same size transcript was detected at much lower levels in the V23-like (ht1/ht1) flowers that contained little or no quercetin (Q-) (lanes 3-8). A reduced level of transcript was also detected in VR leaves (lane 1) and V23 petals (lane 2).

Figures 7a, 7b and 7c Nucleotide sequence and predicted amino acid sequence for the OGR-38 cDNA clone contained in pCGP1805 (Figure 8).

Figure 8 Diagrammatic representation of the plasmid pCGP1805 containing the OGR-38 cDNA clone from *P. hybrida*.

Figure 9 Diagrammatic representation of the yeast expression plasmid pCGP1646. The OGR-38 cDNA insert from pCGP1805 was cloned in a "sense" orientation behind the yeast glyceraldehyde-3-phosphate dehydrogenase promoter (PGAD) in the expression vector pYE22m.

Figure 10 Flavonoid 3'-hydroxylase assay of yeast extracts using [^{3}H]-naringenin as substrate. The representation of an autoradiograph shows conversion of [^{3}H]-naringenin to the 3'-hydroxylated derivative eriodictyol by extracts of yeast transformed with the plasmid pCGP1646 (Lanes 4-7). No 3'-hydroxylase activity was detected in untransformed yeast (Lane 1). The positive controls included extracts of yeast transformed with the plasmid pCGP618 containing the *Hf2* cDNA coding region (as described in the US Patent Number 5,349,125) converting the [^{3}H]-naringenin to eriodictyol and penta-hydroxy flavanone (Lane 2) and pCGP621 containing the 651 cDNA clone (as described in the International Patent Application, having publication number W093/20206) converting the [^{3}H]-naringenin to eriodictyol (Lane 3).

Figure 11 Diagrammatic representation of the binary plasmid pCGP1867. The *Ht1* cDNA insert (OGR-38) from pCGP1805 was cloned in a "sense" orientation behind the Mac promoter of the expression vector pCGP293.

Figure 12 Diagrammatic representation of the plasmid pCGP1807 containing the KC-1 cDNA clone.

Figure 13 Diagrammatic representation of the plasmid pCGP1808 containing the 0.8 kb *KpnI* fragment of KC-1 from pCGP1807 (Figure 12). Sequence data was generated from the 5' end of this fragment using the reverse sequencing primer.

Figure 14 Partial nucleotide sequence and predicted amino acid sequence for the KC-1 cDNA clone contained in pCGP1808 (Figure 13).

Figure 15 Diagrammatic representation of the binary plasmid pCGP1810. The KC-1 cDNA insert from pCGP1807 was cloned in a "sense" orientation behind the Mac

promoter of the expression vector pCGP293.

Figure 16 Representation of an autoradiograph of a Southern blot probed with ^{32}P-labelled fragments of the Am3Ga1 differential display PCR fragment. Each lane contained a 10 μg sample of *EcoR*V-digested genomic DNA isolated from N8 (Eos^+), K16 (eos^-) or plants of an K16 x N8 F_2 population. Hybridizing bands were detected in the genomic DNA from cyanidin-producing plants (cyanidin +) (Lanes 1, 3, 4, 5, 6, 7, 9, 10, 12 and 15). No specific hybridization was observed in the genomic DNA samples from non-cyanidin-producing plants (cyanidin -) (Lanes 2, 8, 11, 13 and 14).

Figure 17 Representation of an autoradiograph of an RNA blot probed with ^{32}P-labelled fragments of the Am3Ga1 differential display PCR fragment. Each lane contained a 10 μg sample of total RNA isolated from the flowers or leaves of plants of an N8 (Eos^+) x K16 (eos^-) F_2 population. A 1.8 kb transcript was detected in the K16 x N8 F_2 flowers that produced cyanidin (cyanidin +) (plants #1, #3, #4, #5 and #8). No transcript was detected in the K16 x N8 F_2 flowers that did not produce cyanidin (cyanidin -) (plants #6, #11, #12) or in a leaf sample (#13L) from an K16 x N8 F_2 plant that produced cyanidin in the flowers.

Figure 18 Diagrammatic representation of the plasmid pCGP246 containing the sdF3'H RACE-clone from *Antirrhinum majus*.

Figures 19 a & 19b Nucleotide sequence and predicted amino acid sequence for the sdF3'H cDNA clone contained in pCGP246.

Figure 20 Diagrammatic representation of the binary plasmid pCGP250. The sdF3'H cDNA insert, containing the nucleotides 1 through to 1711 (Figure 19) from pCGP246, was cloned in a "sense" orientation behind the Mac promoter of the expression vector pCGP293.

Figure 21 Diagrammatic representation of the binary plasmid pCGP231. The sdF3'H cDNA insert, containing the nucleotides 104 through to 1711 (Figure 19) from pCGP246, was cloned in a "sense" orientation behind the Mac promoter of the expression vector pCGP293.

Figures 22 a & 22b Partial nucleotide sequence and predicted amino acid sequence for the arabidopsis putative F3'H cDNA clone.

Figures 23a & 23b Partial nucleotide sequence and predicted amino acid sequence for the rose putative F3'H cDNA clone. Figure 23a shows the sequence data generated from the 5' end of the R4 cDNA clone. Figure 23b shows the sequence data generated from the 3' end of the R4 cDNA clone.

The amino acid abbreviations used throughout the specification are shown below.

Amino acid	3-letter	single-letter
L-alanine	Ala	A
L-arginine	Arg	R
L-asparagine	Asn	N
L-aspartic acid	Asp	D
L-cysteine	Cys	C
L-glutamine	Gln	Q
L-glutamic acid	Glu	E
L-glycine	Gly	G
L-histidine	His	H
L-isoleucine	Ile	I
L-leucine	Leu	L
L-lysine	Lys	K
L-methionine	Met	M
L-phenylalanine	Phe	F
L-proline	Pro	P
L-serine	Ser	S
L-threonine	Thr	T
L-tryptophan	Trp	W
L-tyrosine	Tyr	Y
L-valine	Val	V

The disarmed microorganism *Agrobacterium tumefaciens* strain AGL0 containing the plasmid pCGP1867, pCGP1810 and pCGP231 were deposited with the Australian Government Analytical Laboratories, 1 Suakin Street, Pymble, New South Wales, 2037, Australia on 23 February, 1996 and were given Accession Numbers 96/10967, 96/10968 and 96/10969, respectively.

EXAMPLE 1-Plant Material

Petunia

The *Petunia hybrida* varieties used are presented in Table 2.

Table 2

Plant variety	Properties	Source/Reference
Old Glory Blue (OGB)	F_1 Hybrid	Ball Seed, USA
Old Glory Red (OGR)	F_1 Hybrid	Ball Seed, USA
V23	*An1, An2, An3, An4, An6, An8, An9, An10, ph1, Hf1, Hf2,* ***ht1****, Rt, po, Bl, Fl*	Wallroth et al. (1986) Doodeman et al. (1984)
R51	*An1, An2, An3, an4, An6, An8, An9, An10, An11, Ph1, hf1, hf2,* ***Ht1****, rt, Po, bl, fl*	Wallroth et al. (1986) Doodeman et al. (1984)
VR	V23 x R51 F_1 Hybrid	
SW63	*An1, An2, An3, an4, An6, An8, An9, An10, An11, Ph1, Ph2, Ph5, hf1, hf2,* ***ht1****, ht2, po, mf1, fl*	I.N.R.A., Dijon France
Skr4	*An1, An2, An3, An4, An6, An11, hf1, hf2,* ***ht1****, Ph1, Ph2, Ph5, rt, Po, Mf1, Mf2, fl*	I.N.R.A., Dijon France
Skr4 x SW63	F_1 Hybrid	

Plants were grown in specialised growth rooms with a 14 hour day length at a light intensity of 10,000 lux and a temperature of 22°C to 26°C.

Carnation

Flowers of *Dianthus caryophyllus* cv. Kortina Chanel were obtained from Van Wyk and Son Flower Supply, Victoria.

Dianthus caryophyllus flowers were harvested at developmental stages defined as follows:

Stage 1: Closed bud, petals not visible.
Stage 2: Flower buds opening: tips of petals visible.
Stage 3: Tips of nearly all petals exposed. "Paint-brush stage".
Stage 4: Outer petals at 45° angle to stem.
Stage 5: Flower fully open.

Snapdragon

The *Antirrhinum majus* lines used were derived from the parental lines K16 (eos^-, del^-) and N8 (Eos^+, del^-). K16 is a homozygous recessive mutant lacking F3'H activity while N8 is wild type for F3'H activity. These lines are similar, though not isogenic. Both parental lines and the seed from a selfed (K16 x N8) F_1 plant were obtained from Dr C. Martin (John Innes Centre, Norwich, UK).

Rose

Flowers of *Rosa hybrida* cv. Kardinal were obtained from Van Wyk and Son Flower Supply, Victoria.

Stages of *Rosa hybrida* flower development were defined as follows:

Stage 1: Unpigmented, tightly closed bud (10-12 mm high; 5 mm wide).
Stage 2: Pigmented, tightly closed bud (15 mm high ; 9 mm wide).
Stage 3: Pigmented, closed bud; sepals just beginning to open (20-25 mm high; 13-15 mm wide)

Stage 4: Flower bud beginning to open; petals heavily pigmented; sepals have separated (bud is 25-30 mm high and 18 mm wide).

Stage 5: Sepals completely unfolded; some curling. Petals are heavily pigmented and unfolding (bud is 30-33 mm high and 20 mm wide).

EXAMPLE 2-Bacterial Strains

The *Escherichia coli* strains used were:

DH5α supE44, Δ(lacZYA-ArgF)U169, φ80lacZΔM15, hsdR17 (r_k-, m_k+), recA1, endA1, gyrA96, thi-1, relA1, deoR (Hanahan, 1983 and BRL, 1986).

XL1-Blue MRF'Δ(mcr A)183, Δ(mcrCB-hsdSMR-mrr)173, endA1, supE44, thi-1, recA1, gyrA96, relA1, lac[F' proAB, $lacI_qZ\Delta M15$, Tn10(Tet^r)]^c (Stratagene)

SOLR e14$^-$ (mcrA), Δ(mcrCB-hsdSMR-mrr)171, sbcC, recB, recJ, umuC::Tn5(kan^r), uvrC,lac, gyrA96, thi-1, relA1, [F'proAB, $lacI^qZ\Delta M15$], Su$^-$ (non-suppressing)(Stratagene)

DH10 B(Zip) F$^-$mcrA, Δ(mrr-hsdRMS-mcrBC), φ80d lacZΔM15, ΔlacX74, deoR, recA1, araD139, Δ(ara, leu)7697, galU, galK λ$^-$, rspL, nupG

XL1-Blue supE44, hsdR17 (r_k-,m_k+), recA1, endA1, gyrA96, thi-1, relA1, lac-, [F'proAB, $lacI^q$, lacZΔM15, Tn10(tet^r)] (Bullock *et al*, 1987).

The disarmed *Agrobacterium tumefaciens* strain AGL0 (Lazo *et al.*, 1991) was obtained from R. Ludwig (Department of Biology, University of California, Santa Cruz, USA).

The cloning vector pBluescript was obtained from Stratagene.

Transformation of the *E. coli* strain DH5α-cells was performed according to the method of Inoue *et al.* (1990).

EXAMPLE 3-General methods

^{32}P-Labelling of DNA Probes

DNA fragments (50 to 100 ng) were radioactively labelled with 50 μCi of [α-^{32}P]-dCTP using an oligolabelling kit (Bresatec). Unincorporated [α-^{32}P]-dCTP was removed by chromatography on a Sephadex G-50 (Fine) column.

DNA Sequence Analysis

DNA sequencing was performed using the PRISM™ Ready Reaction Dye Primer Cycle Sequencing Kits from Applied Biosystems. The protocols supplied by the manufacturer were followed. The cycle sequencing reactions were performed using a Perkin Elmer PCR machine (GeneAmp PCR System9600) and run on an automated 373A DNA sequencer (Applied Biosystems).

Homology searches against Genbank, SWISS-PROT and EMBL databases were performed using the FASTA and TFASTA programs (Pearson and Lipman, 1988) or BLAST programs (Altschul *et al.*, 1990).

EXAMPLE 4-Isolation of a flavonoid 3'-hydroxylase (F3'H) cDNA clone corresponding to the Ht1 locus from *P. hybrida*

In order to isolate a cDNA clone that was linked to the *Ht1* locus and that represented the flavonoid 3'-hydroxylase (F3'H) in the petunia flavonoid pathway, a petal cDNA library was prepared from RNA isolated from stages 1 to 3 of Old Glory Red (OGR) petunia flowers. OGR flowers contain cyanidin-based pigments and have high levels of flavonoid 3'-hydroxylase activity. The OGR cDNA library was screened with a mixture of ^{32}P-labelled fragments isolated from three cytochrome P450 cDNA clones known to be involved in the flavonoid pathway and from one cytochrome P450 cDNA clone (651) that had flavonoid 3'-hydroxylase activity in yeast. These included a petunia cDNA clone representing the cinnamate-4-hydroxylase (C4H) and two petunia cDNA clones (coded by the *Hf1* and *Hf2* loci) representing flavonoid 3' 5'-hydroxylase

(F3′ 5′H) (Holton *et al.*, 1993).

Construction of OGR cDNA library

Total RNA was isolated from the petal tissue of *P. hybrida* cv OGR stage 1 to 3 flowers using the method of Turpen and Griffith (1986). Poly(A)$^+$ RNA was selected from the total RNA, using oligotex-dT™ (Qiagen).

A ZAP-cDNA Gigapack III Gold Cloning kit (Stratagene) was used to construct a directional petal cDNA library in λZAP using 5 μg of poly(A)+ RNA isolated from stages 1 to 3 of OGR as template. The total number of recombinants obtained was 2.46×10^6.

After transfecting XL1-Blue MRF′ cells, the packaged cDNA mixture was plated at 50,000 plaques per 15 cm diameter plate. The plates were incubated at 37°C for 8 hours, and the phage were eluted in 100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl pH 8.0, 0.01% (w/v) gelatin (Phage Storage Buffer (PSB)) (Sambrook *et al.*, 1989). Chloroform was added and the phage stored at 4°C as an amplified library.

100,000 pfu of the amplified library were plated onto NZY plates (Sambrook *et al.*, 1989) at a density of 10,000 plaques per 15 cm plate after transfecting XL1-Blue MRF′ cells, and incubated at 37°C for 8 hours. After incubation at 4°C overnight, duplicate lifts were taken onto Colony/Plaque Screen™ filters (DuPont) and treated as recommended by the manufacturer.

Isolation of probes

F3′5′H probes

The two flavonoid 3′, 5′ hydroxylases corresponding to the *Hf1* or *Hf2* loci isolated as described in Holton *et al.* (1993) and US Patent Number 5,349,125, were used in the screening process.

C4H cDNA clones from petunia

A number of cytochrome P450 cDNA clones were isolated in the screening process used to isolate the two flavonoid 3', 5' hydroxylase cDNA clones corresponding to the *Hf1* or *Hf2* loci (Holton *et al.*, 1993; US Patent Number 5,349,125). One of these cDNA clones (F1) (contained in pCGP161) (Figure 2) was identified as representing a cinnamate 4-hydroxylase (C4H), based on sequence identity with a previously-characterised C4H clone from mung bean (Mizutani *et al.*, 1993). Sequence data was generated from 295 nucleotides at the 5' end of the petunia F1 cDNA clone. There was 83.1% similarity with the mung bean C4H clone over the 295 nucleotides sequenced and 93.9% similarity over the predicted amino acid sequence.

651 cDNA clone

The isolation and identification of the 651 cDNA clone contained in pCGP619 (Figure 5) was described in the International Patent Application, having publication number W093/20206. A protein extract of yeast containing the 651 cDNA clone under the control of the yeast glyceraldehyde-3-phosphate dehydrogenase promoter of pYE22m (Tanaka *et al.*, 1988) exhibited F3'H activity.

Screening of OGR Library

Prior to hybridization, the duplicate plaque lifts were washed in prewashing solution (50 mM Tris-HCl pH7.5, 1 M NaCl, 1 mM EDTA, 0.1% (w/v) sarcosine) at 65°C for 30 minutes; stripped in 0.4 M sodium hydroxide at 65°C for 30 minutes; then washed in a solution of 0.2 M Tris-HCl pH 8.0, 0.1 x SSC, 0.1% (w/v) SDS at 65°C for 30 minutes and finally rinsed in 2 x SSC, 1.0% (w/v) SDS.

The lifts from the OGR cDNA library were screened with ^{32}P-labelled fragments of (1) a 0.7 kb *EcoR*I/*Xho*I fragment from pCGP161 containing the C4H cDNA clone (Figure 2), (2) a 1.6 kb *BspH*I/*Fsp*I fragment from pCGP602 containing the *Hf1* cDNA clone (Figure 3), (3) a 1.3 kb *EcoR*I/*Xho*I fragment and a 0.5 kb *Xho*I fragment from pCGP175 containing the coding region of the *Hf2* cDNA clone (Figure 4) and (4) a 1.8 kb *EcoR*I/*Xho*I fragment pCGP619 containing the 651 cDNA clone (Figure 5).

Hybridization conditions included a prehybridization step in 10% (v/v) formamide, 1M NaCl, 10% (w/v) dextran sulphate, 1% (w/v) SDS at 42°C for at least 1 hour. The ^{32}P-labelled fragments (each at 1x10^{6}cpm/mL) were then added to the hybridization solution and hybridization was continued at 42°C for a further 16 hours. The filters were then washed in 2 x SSC, 1% (w/v) SDS at 42°C for 2 x 1 hour and exposed to Kodak XAR film with an intensifying screen at -70°C for 16 hours.

Two hundred and thirty strongly hybridizing plaques were picked into PSB. Of these, 39 were rescreened to isolate purified plaques, using the hybridization conditions as described for the initial screening of the cDNA library. The plasmids contained in the λZAP bacteriophage vector were rescued and sequence data was generated from the 3′ and 5′ ends of the cDNA inserts. Based on sequence homology, 27 of the 39 were identical to the petunia cinnamate 4-hydroxylase cDNA clone, 2 of the 39 were identical to the *Hf1* cDNA clone and 7 of the 39 did not represent cytochrome P450s. The remaining 3 cDNA clones (designated as OGR-27, OGR-38, OGR-39) represented "new" cytochrome P450s, compared to the cytochrome P450 clones used in the screening procedure, and were further characterised.

EXAMPLE 5 -RFLP analysis

There are two genetic loci in *P. hybrida*, *Ht1* and *Ht2*, that control flavonoid 3′-hydroxylase activity (Tabak *et al.*, 1978; Wiering and de Vlaming, 1984). Ht1 is expressed in both the limb and the tube of *P. hybrida* flowers and gives rise to higher levels of F3′H activity than does *Ht2* which is only expressed in the tube. The F3′H is able to convert dihydrokaempferol and naringenin to dihydroquercetin and eriodictyol, respectively. In a flower producing delphinidin-based pigments, F3′H activity is masked by F3′5′H activity. Therefore the F3′H/F3′5′H assay (Stotz and Forkmann, 1982) is useless in determining the presence or absence of F3′H activity. The enzyme flavonol synthase is able to convert dihydrokaempferol to kaempferol and dihydroquercetin to quercetin (Figure 1a). Myricetin, the 3′, 5′ hydroxylated flavonol, is produced at low levels in petunia flowers. Therefore, analysing the flowers for the 3′ hydroxylated flavonol, quercetin, infers the presence of F3′H activity.

Restriction Fragment Length Polymorphism (RFLP) analysis of DNA isolated from individual plants in a VR (Ht1/ht1) x V23 (ht1/ht1) backcross was used to determine which, if any, of the cDNA clones representing P450s were linked to the *Ht1* locus. Northern analysis of RNA isolated from these plants was also used to detect the presence or absence of a transcript in these lines.

Flowers from a VR (Ht1/ht1) x V23 (ht1/ht1) backcross population were analysed for the presence of the flavonols, kaempferol and quercetin. VR (Ht1/ht1) flowers accumulate quercetin and low levels of kaempferol while V23 (ht1/ht1) flowers accumulate kaempferol but little or no quercetin. Individual plants from the VR (Ht1/ht1) x V23 (ht1/ht1) backcross were designated as VR-like (Ht1/ht1), if a substantial level of quercetin was detected in the flower extracts, and V23-like (ht1/ht1), if little or no quercetin but substantial levels of kaempferol were detected in the flower extracts.

Isolation of Genomic DNA

DNA was isolated from leaf tissue essentially as described by Dellaporta *et al.*, (1983). The DNA preparations were further purified by CsCl buoyant density centrifugation (Sambrook *et al.*, 1989).

Southern blots

The genomic DNA (10 μg) was digested for 16 hours with 60 units of *Eco*RI and electrophoresed through a 0.7% (w/v) agarose gel in a running buffer of TAE (40 mM Tris-acetate, 50 mM EDTA). The DNA was then denatured in denaturing solution (1.5 M NaCl/0.5 M NaOH) for 1 to 1.5 hours, neutralized in 0.5 M Tris-HCl (pH 7.5)/ 1.5 M NaCl for 2 to 3 hours and then transferred to a Hybond N (Amersham) filter in 20 x SSC.

Northern blots

Total RNA was isolated from the petal tissue of *P. hybrida* cv OGR stage 1 to 3 flowers using the method of Turpen and Griffith (1986).

RNA samples were electrophoresed through 2.2 M formaldehyde/1.2% (w/v) agarose gels using running buffer containing 40 mM morpholinopropanesulphonic acid (pH 7.0), 5 mM sodium acetate, 0.1 mM EDTA (pH 8.0). The RNA was transferred to Hybond-N filters (Amersham) as described by the manufacturer.

Hybridization and washing conditions

Southern and Northern blots were probed with ^{32}P-labelled cDNA fragment (10^8 cpm/μg, 2 x 10^6 cpm/mL). Prehybridizations (1 hour at 42°C) and hybridizations (16 hours at 42°C) were carried out in 50% (v/v) formamide, 1 M NaCl, 1% (w/v) SDS, 10% (w/v) dextran sulphate. Filters were washed in 2 x SSC, 1% (w/v) SDS at 65°C for 1 to 2 hours and then 0.2 x SSC, 1% (w/v) SDS at 65°C for 0.5 to 1 hour. Filters were exposed to Kodak XAR film with an intensifying screen at -70°C for 16 hours.

RFLP and Northern analysis of the cytochrome P450 fragments

RFLP analysis was used to investigate linkage of the genes corresponding to the OGR-27, OGR-38 and OGR-39 cDNA clones to the *Ht1* locus.

^{32}P-labelled fragments of OGR-27, OGR-38 and OGR-39 cDNA clones were used to probe Northern blots and Southern blots of genomic DNA isolated from individual plants in the VR x V23 backcross population. Analysis of *EcoR*I digested genomic DNA isolated from a VR x V23 backcross population revealed a RFLP for the OGR-38 probe which was linked to *Ht1*. Furthermore there was a much reduced level of transcript detected in the V23-like lines compared to high levels of transcript detected in VR-like lines (Figure 6).

The data provided strong evidence that the OGR-38 cDNA clone corresponded to the *Ht1* locus and represented a F3'H.

EXAMPLE 6-Complete sequence of OGR-38

The complete sequence of the cDNA clone (OGR-38) (Figures 7a, b and c) contained in the plasmid pCGP1805 (Figure 8) was determined by compilation of sequence from different pUC18 subclones obtained using standard procedures for the generation of randomly overlapping clones (Sambrook *et al.*, 1989). The sequence contained an open reading frame of 1536 bases which encodes a putative polypeptide of 512 amino acids.

The nucleotide sequence of OGR-38 was compared with the sequences of the cytochrome P450 probes used in the screening process along with other petunia cytochrome P450 sequences (US Patent Number 5,349,125). The nucleotide sequence of OGR-38 was most similar to the nucleic acid sequence of the flavonoid 3′,5′-hydroxylases representing *Hf1* and *Hf2* loci from *P. hybrida* (Holton *et al.*, 1993). The *Hf1* clone was 58.9% similar at the nucleic acid level over 1471 nucleotides and 49.9% similar at the amino acid level over 513 amino acids, while the *Hf2* clone was 58.9% similar at the nucleic acid level over 1481 nucleotides and 49.1% similar at the amino acid level over 511 amino acids.

EXAMPLE 7-The F3′H assay of the *Ht1* cDNA clone (OGR-38) expressed in yeast

Construction of pCGP1646

The plasmid pCGP1646 (Figure 9) was constructed by cloning the cDNA insert from pCGP1805 (Figure 8) in a "sense" orientation behind the yeast glyceraldehyde-3-phosphate dehydrogenase promoter of pYE22m (Tanaka *et al.*, 1988).

The plasmid pCGP1805 was linearised by digestion with *Asp*718. The overhanging 3′ ends were removed using DNA polymerase (Klenow fragment) according to standard protocols (Sambrook *et al.*, 1989). The 1.8 kb OGR-38 cDNA fragment was released upon digestion with *Sma*I. The cDNA fragment was isolated and purified using the Bresaclean kit (Bresatec) and ligated with blunted *EcoR*I ends of pYE22m. The plasmid pYE22m had been digested with *EcoR*I and the overhanging 5′ ends were removed using DNA polymerase (Klenow fragment) according to standard protocols

(Sambrook *et al.*, 1989). The ligation was carried out with the Amersham Ligation kit using 100ng of the 1.8 kb OGR-38 fragment and 150ng of the prepared yeast vector, pYE22m. Correct insertion of the insert in pYE22m was established by *Xho*I/*Sal*I restriction enzyme analysis of the plasmid DNA isolated from ampicillin resistant transformants.

Yeast transformation

The yeast strain G-1315 (Mat α, trpl) (Ashikari *et al.*, 1989) was transformed with pCGP1646 according to Ito *et al.* (1983). The transformants were selected by their ability to restore G-1315 to tryptophan prototrophy.

Preparation of yeast extracts for assay of F3'H activity

A single isolate of G-1315/pCGP1646 was used to inoculate 50 mL of Modified Burkholder's medium (20g/L dextrose, 2g/L L-asparagine, 1.5g/L KH_2PO_4, 0.5g/L $MgSO_4.7H_2O$, 0.33g/L $CaCl_2$, 2g/L $(NH_4)_2SO_4$, 0.1 mg/L KI, 0.92g/L $(NH_4)_6Mo7O_{24}.4H_2O$, 0.1g/L nitrilotriacetic acid, 0.99 mg/L $FeSO_4.7H_2O$, 1.25 mg/L EDTA, 5.47 mg/L $ZnSO_4.7H_2O$, 2.5 mg/L $FeSO_4.7H_2O$, 0.77 mg/L $MnSO_4.7H_2O$, 0.196 mg/L $CuSO_4.5H_2O$, 0.124 mg/L $Co(NH_4)_2(SO_4)_2.6H_2O$, 0.088 mg/L $Na_2B_4O_7.10H_2O$, 0.2 mg/L thiamine, 0.2 mg/L pyridoxine, 0.2 mg/L nicotinic acid, 0.2 mg/L pantothenate, 0.002 mg/L biotin, 10 mg/L inositol) which was subsequently incubated until the value at OD_{600} was 1.8 at 30°C. Cells were collected by centrifugation and resuspended in Buffer 1 [10 mM Tris-HCl buffer (pH 7.5) containing 2 M sorbitol, 0.1 mM DTT, 0.1 mM EDTA, 0.4 mM phenylmethylsulfonyl fluoride (PMSF) and 5 mg yeast lytic enzyme/mL]. Following incubation for 1 hour at 30°C with gentle shaking, the cells were pelleted by centrifugation and washed in ice cold Buffer 2 [10 mM Tris-HCl (pH7.5) containing 0.65 M sorbitol, 0.1 mM DTT, 0.1 mM EDTA, 0.4 mM PMSF]. The cells were then resuspended in Buffer 2 and sonicated using six 15-second bursts with a Branson Sonifier 250 at duty cycle 30% and output control 10%. The sonicated suspension was centrifuged at 10,000 rpm for 30 minutes and the supernatant was centrifuged at 13,000 rpm for 90 minutes. The microsomal pellet was resuspended in assay buffer (100 mM potassium phosphate (pH 8), 1 mM EDTA, 20 mM 2-mercaptoethanol) and 100 μL was assayed for

activity.

F3'H Assay

F3'H enzyme activity was measured using a modified version of the method described by Stotz and Forkmann (1982). The assay reaction mixture typically contained 100 μL of yeast extract, 5 μL of 50 mM NADPH in assay buffer (100 mM potassium phosphate (pH8.0), 1 mM EDTA and 20 mM 2-mercaptoethanol) and 10 μCi of [^{3}H]-naringenin and was made up to a final volume of 210 μL with the assay buffer. Following incubation at 23°C for 2-16 hours, the reaction mixture was extracted with 0.5 mL of ethylacetate. The ethylacetate phase was dried under vacuum and then resuspended in 10 μL of ethylacetate. The tritiated flavonoid molecules were separated on cellulose thin layer plates (Merck Art 5577, Germany) using a chloroform: acetic acid: water (10:9:1 v/v) solvent system. The reaction products were localised by autoradiography and identified by comparison to non-radioactive naringenin and eriodictyol standards which were run alongside the reaction products and visualised under UV light.

F3'H activity was detected in extracts of G1315/pCGP1646, but not in extracts of non-transgenic yeast (Figure 10). From this it was concluded that the cDNA insert from pCGP1805 (OGR-38), which was linked to the *Ht1* locus, encoded a F3'H.

EXAMPLE 8-Transient expression of the *Ht1* cDNA clone (OGR-38) in plants

Construction of pCGP1867

Plasmid pCGP1867 (Figure 11) was constructed by cloning the cDNA insert from pCGP1805 in a "sense" orientation behind the Mac promoter (Comai *et al.*, 1990) of pCGP293 (Brugliera *et al.*, 1994). The plasmid pCGP1805 was digested with *Xba*I and *Kpn*I to release the cDNA insert. The cDNA fragment was isolated and purified using the Bresaclean kit (Bresatec) and ligated with *Xba*I/*Kpn*I ends of the pCGP293 binary vector. The ligation was carried out using the Amersham ligation kit. Correct insertion of the fragment in pCGP1867 was established by *Xba*I/*Kpn*I restriction enzyme analysis of DNA isolated from gentamycin-resistant transformants.

Transient expression of the *Ht1* cDNA clone (OGR-38) in petunia petals

In order to rapidly determine whether the OGR-38 cDNA fragment in pCGP1867 represented a functional F3'H in plants, a transient expression study was established. Petals of the mutant *P. hybrida* line Skr4 x SW63 were bombarded with gold particles (1μm diameter) coated with pCGP1867 DNA.

Gold particles were prewashed 3 times in 100% ethanol and resuspended in sterile water. For each shot, 1 μg of pCGP1867 DNA, 0.5 mg of gold particles, 10 μL of 2.5 M $CaCl_2$ and 2 μL of 100 mM spermidine (free base) were mixed by vortexing for 2 minutes. The DNA coated gold particles were pelleted by centrifugation, washed twice with 100% ethanol and finally resuspended in 10 μL of 100% ethanol. The suspension was placed directly on the centre of the macrocarrier and allowed to dry.

Stages 1 and 2 of Skr4 x SW63 flowers were cut vertically into halves and partially embedded in MS solid media (3% (w/v) sucrose, 100 mg/L myo-inositol, 1xMS salts, 0.5 mg/L pyridoxine-HCl, 0.1 mg/L thiamine-HCl, 0.5 mg/L nicotinic acid and 2 mg/L glycine). The petals were placed so that the inside of the flower buds were facing upwards. A Biolistic PDS-1000/He System (Bio-Rad), using a Helium gas pressure of 900psi and a chamber vacuum of 28 inches of mercury, was used to project the gold particles into the petal tissue. After 6-12 hours under lights in a controlled plant growth room at 22°C, red anthocyanin spots were observed on the upper epidermal layer of the petal tissue bombarded with pCGP1867 coated particles. No coloured spots were observed in control petal bombarded with gold particles alone. These results indicated that the OGR-38 cDNA clone under the control of the Mac promoter was functional, at least transiently, in petal tissue.

EXAMPLE 9-Stable expression of the *Ht1* cDNA clone (OGR-38) in petunia petals- Complementation of a *ht1/ht1* petunia cultivar

***A. tumefaciens* transformations**

The plasmid pCGP1867 (Figure 11) was introduced into the *Agrobacterium tumefaciens* strain AGL0 by adding 5 μg of plasmid DNA to 100 μL of competent AGL0 cells prepared by inoculating a 50 mL MG/L (Garfinkel and Nester, 1980) culture and growing for 16 hours with shaking at 28°C. The cells were then pelleted and resuspended in 0.5 mL of 85% (v/v) 100 mM $CaCl_2$/15% (v/v) glycerol. The DNA-*Agrobacterium* mixture was frozen by incubation in liquid N_2 for 2 minutes and then allowed to thaw by incubation at 37°C for 5 minutes. The DNA/bacterial mix was then placed on ice for a further 10 minutes. The cells were then mixed with 1 mL of LB (Sambrook *et al.*, 1989) media and incubated with shaking for 16 hours at 28°C. Cells of *A. tumefaciens* carrying pCGP1867 were selected on LB agar plates containing 10 μg/mL gentamycin. The presence of pCGP1867 was confirmed by Southern analysis of DNA isolated from the gentamycin-resistant transformants.

Petunia transformations

(a) Plant Material

Leaf tissue from mature plants of *P. hybrida* cv Skr4 x SW63 was treated in 1.25% (w/v) sodium hypochlorite for 2 minutes and then rinsed three times in sterile water. The leaf tissue was then cut into 25 mm^2 squares and precultured on MS media (Murashige and Skoog, 1962) supplemented with 0.05 mg/L kinetin and 1.0 mg/L 2,4-dichlorophenoxyacetic acid (2,4-D) for 24 hours.

(b) Co-cultivation of *Agrobacterium* and Petunia Tissue

A. tumefaciens strain AGL0 containing the binary vector pCGP1867 (Figure 11) was maintained at 4°C on MG/L agar plates with 100 mg/L gentamycin. A single colony was grown overnight in liquid medium containing 1% (w/v) Bacto-peptone, 0.5% (w/v) Bacto-yeast extract and 1% (w/v) NaCl. A final concentration of 5 x 10^8 cells/mL was prepared the next day by dilution in liquid MS medium containing B5 vitamins (Gamborg *et al.*, 1968) and 3% (w/v) sucrose (BPM). The leaf discs were

dipped for 2 minutes into BPM containing AGL0/pCGP1867. The leaf discs were then blotted dry and placed on co-cultivation media for 4 days. The co-cultivation medium consisted of SH medium (Schenk and Hildebrandt, 1972) supplemented with 0.05 mg/L kinetin and 1.0 mg/L 2,4-D and included a feeder layer of tobacco cell suspension spread over the co-cultivation medium with a filter paper placed on top of the tobacco cell suspension.

(c) Recovery of transgenic petunia plants

After co-cultivation, the leaf discs were transferred to selection medium (MS medium supplemented with 3% (w/v) sucrose, 2 mg/L α-benzylaminopurine (BAP), 0.5 mg/L α-naphthalene acetic acid (NAA), 300 mg/L kanamycin, 350 mg/L cefotaxime and 0.3% (w/v) Gelrite Gellan Gum (Schweizerhall)). Regenerating explants were transferred to fresh selection medium after 4 weeks. Adventitious shoots which survived the kanamycin selection were isolated and transferred to BPM containing 100 mg/L kanamycin and 200 mg/L cefotaxime for root induction. All cultures were maintained under a 16 hour photoperiod (60 μmol. m-2, s-1 cool white fluorescent light) at 23$\pm$ 2°C. When roots reached 2-3 cm in length the transgenic petunia plantlets were transferred to autoclaved Debco 51410/2 potting mix in 8 cm tubes. After 4 weeks plants were replanted into 15 cm pots using the same potting mix and maintained at 23°C under a 14 hour photoperiod (300 μmol. m-2, s-1 mercury halide light).

EXAMPLE 10-Transgenic plant phenotype analysis
pCGP1867 in Skr4 x SW63

Table 3 shows the various petal and pollen colour phenotypes obtained with Skr4 x SW63 plants transformed with the pCGP1867 plasmid. The transgenic plants #593A, 590A, 571A, 589A, 592A and 591A produced flowers with altered petal colour. The anthers and pollen of the flowers from plants #593A, 590A, 589A, 592A and 591A were pink as compared to those of the control Skr4 x SW63 plant, which were white. The changes in anther and pollen colour that were observed on introduction of plasmid pCGP1867 into Skr4 x SW63 petunia plants was an unanticipated outcome. The codes are taken from the Royal Horticultural Society's Colour Chart (RHSCC). They

provide an alternative means by which to describe the colour phenotypes observed. The designated numbers, however, should be taken only as a guide to the perceived colours and should not be regarded as limiting the possible colours which may be obtained.

Table 3 Summary of petal, anther and pollen colours obtained in Skr4 xSW63 plants transformed with pCGP1867

Accession Number	Petal Limb Colour	RHSCC Code (petal limb)	Anther & Pollen Colour
Skr4 x SW63 control (594A)	very pale lilac	69B/73D	white
593A	dark pink	67B	pink
590A	dark pink and pink sectors	sectored 67B and 73A	pink
571A	pink	68A and B	pink
589A	dark pink	68A	pink
592A	pink and light pink sectors	68A and 68B	light pink
591A	dark pink	68A	pink
570A	very pale lilac	69B/73D	white

The expression of the introduced *Ht1* cDNA in the Skr4 x SW63 hybrid had a marked effect on flower colour. The stamen tissue of the non transgenic control is white whereas the same tissue in most of the transgenic plants was pink. In addition, expression of the *Ht1* cDNA in the Skr4 x SW63 hybrid conferred a dark pink hue to the corolla which is normally very pale lilac.

EXAMPLE 11-Analysis of products

The anthocyanidins and flavonols produced in the petals and stamens (included the pollen, anthers and filaments) of the Skr4 x SW63 plants transformed with pCGP1867 were analysed by thin layer chromatography.

Extraction of anthocyanins and flavonols

Prior to TLC analysis, the anthocyanin and flavonol molecules present in petal and stamen extracts were acid hydrolysed to remove glycosyl moieties from the anthocyanidin or flavonol core. Anthocyanidin and flavonol standards were used to help identify the compounds present in the floral extracts.

Anthocyanins and flavonols were extracted and hydrolysed by boiling between 100 to 200 mg of petal limbs, or five stamens, in 1 mL of 2 M hydrochloric acid for 30 minutes. The hydrolysed anthocyanins and flavonols were extracted with 200 μL of iso-amylalcohol. This mixture was then dried down under vacuum and resuspended in a smaller volume of methanol/1% (v/v) HCl. The volume of methanol/1% (v/v) HCl used was based on the initial fresh weight of the petal so that the relative levels of flavonoids in the petals could be estimated. Extracts from the stamens were resuspended in 1 μL of methanol/1% (v/v) HCl. A 1 μL aliquot of the extracts from the pCGP1867 in Skr4 x SW63 petals and stamens was spotted onto a TLC plate.

TLC analysis of floral extracts

Acid-hydrolysed floral extracts were run in a Forestal solvent system (HOAc:water:HCl; 30: 10: 3) (Markham, 1982). Table 4 shows the results of the TLC analysis of the anthocyanidins and flavonols present in some of the flowers and stamens of the transgenic Skr4 x SW63 petunia plants transformed with pCGP1867. Indicative relative amounts of the flavonols and anthocyanidins (designated with a "+" to "+++") were estimated by comparing the intensities of the spots observed on the TLC plate.

Table 4 Relative levels of anthocyanidins and flavonols detected in the petal limbs and stamens of Skr4 x SW63 plants transformed with pCGP1867.

Acc#	Petal Colour	Anthocyanidins			Flavonols	
		Malvidin	Cyanidin	Peonidin	Kaempferol	Quercetin
Skr4 x SW63 control petal limb	pale lilac	+/-	-	-	+	-
593A petal limb	dark pink	-	+	+++	-	++
571A petal limb	pink	-	+	+	-	+
589A petal limb	dark pink	-	+	++	-	++
570A petal limb	pale lilac	+/-	-	-	+	-
Skr4 x SW63 control stamens	white	-	-	-	+++	+
					-	
593A stamens	pink	-	-	++	-	+++

Introduction of the *Ht1* cDNA into Skr4 x SW63 led to production of the 3'-hydroxylated flavonoids, quercetin, peonidin and some cyanidin in the petals. Peonidin is the methylated derivative of cyanidin (Figures 1a and 1b). The non-transgenic Skr4 x SW63 control produced only kaempferol and a small amount of

malvidin (Table 4). Although Skr4 x SW63 is homozygous recessive for both the *Hf1* and *Hf2* genes these mutations do not completely block production of F3'5'H (see US Patent Number 5,349,125) and low levels of malvidin are produced to give the petal limb a pale lilac colour.

The stamens with the pink pollen and anthers produced by the transgenic plant #593A contained peonidin and quercetin, while the non-transgenic Skr4 x SW63 control with white pollen and anthers contained kaempferol and a low level of quercetin (Table 4).

The accumulation of the 3'-hydroxylated anthocyanidin, peonidin, in the petals and stamens of the transgenic Skr4 x SW63/pCGP1867 plants correlated with the pink and dark pink colours observed in the petals, anthers and pollen of the same plants.

EXAMPLE 12-Isolation of putative F3'H cDNA clone from *Dianthus caryophyllus*

In order to isolate a carnation F3'H cDNA clone, the petunia Ht1 linked, F3'H cDNA clone (OGR-38) contained in pCGP1805 (described above) was used to screen a Kortina Chanel petal cDNA library under low stringency conditions.

Construction of Kortina Chanel cDNA library

Twenty micrograms of total RNA isolated (as described previously) from stages 1, 2 and 3 of Kortina Chanel flowers was reverse transcribed in a 50 μL volume containing 1 x Superscript™ reaction buffer, 10 mM dithiothreitol (DTT), 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 500 μM 5-methyl-dCTP, 2.8 μg Primer-Linker oligo from ZAP-cDNA Gigapack III Gold cloning kit (Stratagene) and 2 μL Superscript™ reverse transcriptase (BRL). The reaction mix was incubated at 37°C for 60 minutes, then placed on ice. A ZAP-cDNA Gigapack III Gold Cloning kit (Stratagene) was used to complete the library construction. The total number of recombinants was 2.4×10^6.

A total of 200,000 pfu of the packaged cDNA was plated at 10,000 plaques per 15 cm diameter plate after transfecting XL1-Blue MRF' cells. The plates were incubated at 37°C for 8 hours, then stored overnight at 4°C. Duplicate lifts were taken onto

Colony/Plaque Screen™ filters (DuPont) and treated as recommended by the manufacturer.

Screening of Kortina Chanel petal cDNA library for a F3'H cDNA clone

Prior to hybridization, the duplicate plaque lifts were treated as described previously. The duplicate lifts from the Kortina Chanel petal cDNA library were screened with ^{32}P-labelled fragments of the 1.8 kb *EcoRI/Xho*I insert from pCGP1805 (Figure 8). Low stringency conditions, as described for the screening of the petunia OGR cDNA library were used.

One strongly hybridizing plaque was picked into PSB, rescreened as detailed above to isolate purified plaques and was characterized by sequence analysis.

The KC-1 cDNA insert contained in pCGP1807 (Figure 12) was released upon digestion with *EcoRI/Xho*I and is approximately 2 kb in length. Readable sequence data could not be generated from the 3' end, so a 800 bp *Kpn*I fragment covering the 3' region of KC-1 was subcloned into pBluescript to give pCGP1808 (Figure 13). Sequence data covering 458 nucleotides was generated from the 5' end of this fragment to give the sequence detailed in Figure 14.

The sequence of the KC-1 cDNA clone showed 68.8% similarity over 365 nucleotides and 74.1% similarity over 116 amino acids to that of the petunia OGR-38 F3'H cDNA clone .

Construction of pCGP1810

Plasmid pCGP1810 (Figure 15) was constructed by cloning the cDNA insert from pCGP1807 in a "sense" orientation behind the Mac promoter (Comai *et al.*, 1990) of pCGP90 (US Patent Number 5,349,125), a pCGP293 based construct (Brugliera *et al.*, 1994). The plasmid pCGP1805 was digested with *Bam*HI and *Apa*I to release the KC-1 cDNA insert. The cDNA fragment was isolated and purified using the Bresaclean kit (Bresatec). The pCGP90 binary vector was digested with *Bam*HI and *Apa*I to separate the linearised vector and the *Hf1* cDNA insert. The linearised vector was

isolated and purified using the Bresaclean kit (Bresatec) and ligated with *Bam*HI/*Apa*I ends of the KC-1 cDNA clone. The ligation was carried out using the Amersham ligation kit. Ligation of the insert in pCGP1810 was verified by *Bam*HI/*Apa*I restriction enzyme analysis of DNA isolated from gentamycin-resistant transformants.

The binary vector pCGP1810 was introduced into *A. tumefaciens* strain AGL0 cells. The pCGP1810/AGL0 cells have been introduced into Skr4 x SW63 petunia plants, to test for stable expression and activity of the enzyme encoded by the gene corresponding to the KC-1 cDNA clone.

EXAMPLE 13-Isolation of putative F3'H cDNA clone from *Antirrhinum majus* using a differential display approach

A novel approach was employed to isolate a cDNA sequence encoding F3'H from *Antirrhinum majus* (snapdragon). Modified methods based on the protocols for (i) isolation of plant cytochrome P450 sequences using redundant oligonucleotides (Holton *et al.* 1993) and (ii) differential display of eukaryotic messenger RNA (Liang and Pardee, 1992) were combined, to compare the petal cytochrome P450 transcript populations between wild type and F3'H mutant snapdragon lines. Direct cloning of differentially expressed cDNA fragments allowed their further characterisation by Northern, RFLP and sequence analysis to identify putative F3'H encoding sequences. A full-length cDNA was obtained using the RACE protocol of Frohman *et al.* (1988) and the clone was shown to encode a functional F3'H following transient expression in petunia petal cells.

Plant Material

The *Antirrhinum majus* lines used were derived from the parental lines K16 (eos^-, del^-) and N8 (Eos^+, del^-). K16 is a homozygous recessive mutant lacking F3'H activity, while N8 is wild type for F3'H activity. These lines are similar, though not isogenic. The seed of capsule $E228^2$ from the selfed K16 x N8 F_1 plant (#E228) was germinated and the resultant plants (K16 x N8 F_2 plants) were scored for the presence or absence of cyanidin, a product of F3'H activity. The presence of

cyanidin could be scored visually as the flowers were a crimson colour, unlike the mutant plants which were a pink colour (from pelargonidin-derived pigments). The accuracy of the visual scoring was confirmed by TLC analysis of the petal anthocyanins as described in Example 11.

Of 13 plants raised from the E228[2] seed, 9 (#1, #2, #3, #4, #5, #7, #8, #10, #13) produced flowers with cyanidin (Eos^+/Eos^+ and Eos^+/eos^-) while 4 (#6, #9, #11, #12, #14) produced only pelargonidin-derived pigments (eos^-/eos^-).

Synthesis of cDNA

Total RNA was isolated from the leaves and petal tissue of the *A. majus* K16 x N8 F_2 segregating population (E228[2]) using the method of Turpen and Griffith (1986). Contaminating DNA was removed by treating 50 μg total RNA with 1 unit RQ1 RNase-free DNase (Promega) in the presence of 40 units RNasin® ribonuclease inhibitor (Promega) for 3 hours at 37°C in a buffer recommended by the manufacturers. The RNA was then further purified by extraction with phenol/chloroform/iso-amyl alcohol (25:24:1) and subsequent ethanol precipitation.

Anchored poly(T) oligonucleotides, complementary to the upstream region of the polyadenylation sequence, were used to prime cDNA synthesis from *A. majus* petal and leaf RNA. The oligonucleotide sequences (written 5'-3') synthesized were:

polyT-anchA	TTTTTTTTTTTTTTTTA
polyT-anchC	TTTTTTTTTTTTTTTTC
polyT-anchG	TTTTTTTTTTTTTTTTG

Two micrograms of total RNA and 100 pmol of the appropriate priming oligonucleotide were heated to 70°C for 10 minutes, then chilled on ice. The RNA/primer hybrids were then added to a reaction containing 20 units RNasin® (Promega), 25 nM each dNTP, 10 mM DTT and 1x Superscript buffer (BRL). This reaction was heated at 37°C for 2 minutes, then 200 units of Superscript™ reverse transcriptase (BRL) were added and the reaction allowed to proceed for 75 minutes,

after which the reverse transcriptase was inactivated by heating the mixture at 95°C for 20 minutes.

Amplification of cytochrome P450 sequences using PCR

Cytochrome P450 sequences were amplified using redundant oligonucleotides (designed to be complementary to conserved regions near the 3′ end of plant cytochrome P450 coding sequences) and polyT anchored oligonucleotides. A similar approach was previously used to generate cytochrome P450 sequences from *Petunia hybrida* and is described in US Patent Number 5,349,125,

Four oligonucleotides (referred to as upstream primers) derived from conserved regions in plant cytochrome P450 sequences were synthesized. These were (written 5′ to 3′):

WAIGRDP	TGG GCI ATI GGI (A/C)GI GA(T/C) CC
FRPERF	AGG AAT T(T/C)(A/C) GIC CIG A(A/G)(A/C) GIT T
PETHAEM-NEW	CCI TT(T/C) GGI GCI GGI (A/C)GI (A/C)GI ATI TG(T/G) (C/G)CI GG
EFXPERF	GAI TT(T/C) III CCI GAI (A/C)GI TT

The upstream primers were used with each of the polyT anchored oligonucleotides to generate cytochrome P450 sequences in polymerase chain reactions using cDNA as a template. Fifty picomoles of each oligonucleotide was combined with 2 μM of each dNTP, 1.5 mM $MgCl_2$, 1x PCR buffer (Perkin Elmer), 5 μCi α-[^{33}P] dATP (Bresatec, 1500 Ci/mmol), 2.5 units AmpliTaq® DNA polymerase (Perkin Elmer) and 1/10th of the polyT-anchor primed cDNA reaction (from above). Reaction mixes (50 μL) were cycled 40 times between 94°C for 15 seconds, 42°C for 15 seconds, and 70°C for 45 seconds, following an initial 2 minute denaturation step at 94°C. Cycling reactions were performed using a Perkin Elmer 9600 Gene Amp Thermal Cycler.

DNA sequences were amplified using each upstream primer/anchored primer combination and the appropriately primed cDNA template. Each primer combination was used with the cDNA from the petals of the E228[2] plants #3 and #5 (cyanidin-producing flowers) and #12 (non-cyanidin producing flowers). Reactions incorporating leaf cDNA from plant #13 (cyanidin-producing flowers) were also included as negative controls, as F3'H activity is not present at a significant level in healthy, unstressed leaf tissues.

Differential display of cytochrome P450 sequences

^{33}P-labelled PCR fragments were visualised following separation on a 5% (w/v) polyacrylamide/urea denaturing gel (Sambrook *et al.* 1989). A ^{33}P-labelled M13mp18 sequencing ladder was included on the gel to serve as a size marker. The sequencing gel was dried onto Whatman 3MM paper and exposed to Kodak XAR film at room temperature.

Comparison of bands between cyanidin-producing petal samples and the non-cyanidin petal sample revealed 11 bands which represent mRNAs exclusively present in the cyanidin-producing petals. Of these 11 bands, only two were also present (at a reduced intensity) in the leaf sample.

Isolation and cloning of PCR fragments from sequencing gel

PCR products were purified from the dried sequencing gel and reamplified by the method described by Liang *et al.* (1993) using the appropriate primer combination. Amplified cDNAs were purified, following electrophoretic separation on a 1.2% (w/v) agarose/TAE gel, using a Bresaclean kit (Bresatec). The purified fragments were then directly ligated into either commercially prepared pCR-Script™ vector (Stratagene) or *Eco*RV-linearised pBluescript® (Stratagene) which had been T-tailed using the protocol of Marchuk *et al.* (1990).

Sequence of F3'H PCR products

Each of the eleven cloned differential display PCR products (with inserts not exceeding 500 bp) was sequenced on both strands and compared to other known cytochrome P450 sequences involved in anthocyanin biosynthesis, using the FASTA algorithm of Pearson and Lipman (1988).

Of the eleven cDNAs cloned, two, Am1Gb and Am3Ga, displayed strong homology with the petunia OGR-38 F3'H sequence (Examples 4 to 11) and the F3'5'H sequences (Holton *et al.*, 1993). Conserved sequences between clones Am1Gb and Am3Ga suggested that they represent overlapping fragments of the same mRNA. Clone Am3Ga extends from the sequence encoding the haem-binding region of the molecule (as recognised by the "PETHAEM-NEW" oligonucleotide) to the polyadenylation sequence. Clone Am1Gb extends from the cytochrome P450 sequence encoding the conserved "WAIGRDP" amino acid motif (complementary to primer 1) to an area in the 3' untranslated region which was spuriously recognised, again, by the primer 1 ("WAIGRDP") oligonucleotide.

EXAMPLE 14-RFLP analysis of cytochrome P450 cDNAs

Restriction fragment length polymorphism (RFLP) analysis was again used to investigate linkage of the gene corresponding to cDNA clone Am3Ga to the presence, or absence, of cyanidin-producing activity in petals. A ^{32}P-labelled insert of Am3Ga was used to probe Southern blots of genomic DNA isolated from K16 x N8 F_2 segregating plants as well as the parental K16 and N8 lines. Analysis of *Eco*RV-digested genomic DNA from 13 plants of the K16 x N8 F_2 segregating population revealed hybridization only to the sequences of N8 and the K16 x N8 F_2 segregating lines which displayed floral cyanidin production (Figure 16). The K16 x N8 F_2 plants which produced only pelargonidin-derived pigments in their petals (including parental line, K16) showed no specific hybridization. These data indicate a possible deletion of the genomic sequences corresponding to Am3Ga in the mutant K16 plant and, therefore, at least a partial deletion of the F3'H gene in this line.

EXAMPLE 15-Northern analysis of cytochrome P450 cDNAs

Northern analysis was used to confirm the expression profiles of the putative cytochrome P450 fragments as shown by differential display. Ten micrograms of total petal RNA from eight of the K16 x N8 F_2 segregating population was separated on a 1.2% (w/v) agarose/formaldehyde gel (Sambrook *et al.* 1989) and transferred to HybondN nylon membrane (Amersham). Leaf RNA from the cyanidin-producing plant #13 was also included as a negative control in the Northern analysis. ^{32}P-labelled cDNA inserts from clones Am1Gb, Am3Ga and Am3Gb were used to probe the RNA blots.

cDNA probes Am1Gb and Am3Ga both recognised an approximately 1.8 kb transcript which was only detectable in the petals of cyanidin-producing plants (plants #1, #3, #4, #5, #8). No transcript was detectable in the pelargonidin-producing petals (plants #6, #11, #12) or in the leaf sample from plant #13 (Figure 17).

These data, taken with those of the RFLP analysis, provide strong evidence that clones Am1Gb and Am3Ga represent a cytochrome P450 gene which is responsible for F3'H activity in snapdragon. The total lack of a detectable transcript in the petals of non-cyanidin-producing lines confirms the findings of the RFLP analysis, that the loss of cyanidin-producing activity in the K16 line (and the homozygous recessive plants of the K16 x N8 F_2 segregating population) is the result of a deletion in the F3'H structural gene.

EXAMPLE 16-Isolation of a full-length snapdragon F3'H cDNA

The Rapid Amplification of cDNA Ends (RACE) protocol of Frohman *et al.* (1988) was employed to isolate a full-length F3'H cDNA clone using sequence knowledge of the partial Am3Ga clone. A gene-specific primer ("SnapredRace A" -complementary to Am3Ga sequences 361 to 334) was synthesized to allow reverse transcription from petal RNA. A 3' amplification primer ("SnapredRace C" -complementary to Am3Ga (3'UTR) sequences 283 to 259) was also synthesized to bind just upstream of "SnapredRace A". A "poly(C)" primer was used to amplify sequences from the 5' end

of the cDNA molecule.

The sequences of the oligonucleotides used were (written 5'-3'):

Snapred Race A	CCA CAC GAG TAG TTT TGG CAT TTG ACC C
Snapred Race C	GTC TTG GAC ATC ACA CTT CAA TCT G
PolyC race	CCG AAT TCC CCC CCC CC

"Snapred Race A-primed" petal cDNA was poly(G)-tailed and a 5' cDNA fragment amplified with primers "Snapred Race C" and "PolyC race" using the method of Frohman *et al.* (1988). Pfu DNA polymerase (0.15 unit) (Stratagene) was combined with 2.5 units AmpliTaq® DNA polymerase (Perkin Elmer) to increase the fidelity of the PCR reaction.

The resultant 1.71 kb DNA fragment (sdF3'H) was cloned directly into *EcoR*V-linearised pBluescript™ (Stratagene) vector which had been T-tailed using the protocol of Marchuk *et al.* (1990). This plasmid was named pCGP246 (Figure 18).

EXAMPLE 17-Complete sequence of snapdragon F3'H

Convenient restriction sites within the sdF3'H cDNA sequence of pCGP246 were exploited to generate a series of short overlapping subclones in the plasmid vector pUC19. The sequence of each of these subclones was compiled to provide the sequence of the entire sdF3'H RACE cDNA. The sdF3'H cDNA sequence was coupled with that from clone Am3Ga to provide the entire sequence of a snapdragon F3'H cDNA shown in Figure 19 a and b. It contains an open reading frame of 1711 bases which encodes a putative polypeptide of 512 amino acids.

The nucleotide sequence of sdF3'H was compared with the sequences of OGR-38 (Examples 4 to 11), *Hf1*, *Hf2*, and other petunia cytochrome P450 sequences isolated previously (US Patent Number 5,349,125). The nucleotide sequence of sdF3'H was most similar to the nucleic acid sequence of the F3'H cDNA clone (OGR-38) representing the *Ht1* locus from *P. hybrida* having 69% similarity at the nucleic acid

level over 1573 nucleotides and 72.2% similarity at the amino acid level over 507 amino acids.

The *Hf1* clone was 57.3% similar at the nucleic acid level, over 1563 nucleotides and 49.3% similar at the amino acid level, over 491 amino acids, while the *Hf2* clone was 57.7% similar at the nucleic acid level, over 1488 nucleotides and 47.8% similar at the amino acid level, over 508 amino acids.

EXAMPLE 18-Transient expression of sdF3'H in plants

Construction of pCGP250

Plasmid pCGP250 (Figure 20) was created by cloning the entire sdF3'H RACE cDNA insert (from position 1 to 1711 (Figure 19a)) from pCGP246 in the "sense" orientation behind the Mac promoter (Comai *et al.*, 1990) of pCGP293 (Brugliera *et al.*, 1994). The plasmid pCGP246 was digested with *EcoR*I to release the cDNA insert. The cDNA fragment was blunt-ended by repairing the overhangs with the Klenow fragment of DNA polymerase I (Sambrook *et al.*, 1989) and purified, following agarose gel electrophoresis, using a Bresaclean kit (Bresatec). The blunt cDNA fragment was then ligated into the binary vector pCGP293, which had been linearised with *Xba*I and blunt-ended using the Klenow fragment of DNA polymerase I. The ligation was carried out using the Amersham ligation kit. Correct insertion of the insert in pCGP250 was established by *Bam*HI and *Pst*I restriction enzyme analysis of DNA isolated from gentamycin-resistant transformants.

Construction of pCGP231

Plasmid pCGP231 (Figure 21) was created by cloning the RACE cDNA insert from pCGP246, downstream of the first "in-frame" ATG codon (from position 104 to 1711 (Figure 19a)), in the "sense" orientation behind the Mac promoter (Comai *et al.*, 1990) of pCGP293 (Brugliera *et al.*, 1994). The plasmid pCGP246 was digested with *Ssp*I (which recognises a site between the candidate ATG codons) and *Sma*I (with a site in the vector polylinker sequence) to release a blunt-ended cDNA fragment which includes the entire coding region downstream from the second putative initiation codon. The cDNA fragment was then ligated into the binary vector pCGP293, which

had been linearised with *Xba*I and blunt-ended using the Klenow fragment of DNA polymerase I. The ligation was carried out using the Amersham ligation kit. Correct insertion of the insert in pCGP231 was established by *Bam*HI and *Pst*I restriction enzyme analysis of DNA isolated from gentamycin-resistant transformants.

Transient Expression Studies

To determine rapidly whether the pCGP246 sequences in pCGP231 and pCGP250 encoded active flavonoid 3'-hydroxylases in plants, a transient expression study was undertaken. Petals of the mutant *P. hybrida* line Skr4 X SW63 were bombarded with gold particles (1μm diameter) coated with either pCGP231 or pCGP250 plasmid DNA (as described in Example 8).

After 6-12 hours under lights in a controlled plant growth room at 22°C, red anthocyanin spots were observed on the surface of the petal tissue bombarded with pCGP231 coated particles. No coloured spots were observed in petals bombarded with pCGP250 or control petals bombarded with gold particles alone. These results indicated that the pCGP246 coding region (starting at the second ATG, position 121 (Figure 19a)), under the control of the Mac promoter, was functional in petal tissue.

EXAMPLE 19-Isolation of putative F3'H cDNA clone from *Arabidopsis thaliana* using a PCR approach

In order to isolate a cDNA clone representing flavonoid 3'-hydroxylase from *Arabidopsis thaliana*, PCR fragments were generated using primers from the conserved regions of cytochrome P450s. One PCR product (p58092.13) was found to have high sequence similarity with the petunia OGR-38 and snapdragon F3'H cDNA clones. The PCR fragment was then used, together with the *Ht1* cDNA insert (OGR-38) from pCGP1805 to screen an *A. thaliana* cDNA library.

Design of oligonucleotides

Degenerate oligonucleotides for PCR DNA amplification were designed from the consensus amino acid sequence of *Petunia hybrida* cytochrome P450 partial sequences situated near the haem binding domain. Primer degeneracy was established by the

inclusion of deoxyinosine (I) in the third base of each codon (inosine base pairs with similar efficiency to A, T, G, and C), and the inclusion of alternate bases where the consensus sequences were non-specific. Thus the amino-terminal directional primer "Pet Haem" (Petunia haem binding domain), containing the cysteine residue codon crucial for haem binding, and the upstream primer "WAIGRDP" were designed.

WAIGRDP	TGG GCI ATI GGI (A/C)GI GA(T/C) CC
Pet Haem	CCI GG(A/G) CAI ATI C(G/T)(C/T) (C/T)TI CCI GCI CC(A/G) AAI GG

Generation of cytochrome P450 sequences using PCR

Genomic DNA was isolated from *A. thaliana* cv. Columbia using the method described by Dellaporta *et al.* (1987). Polymerase chain reactions for amplification of cytochrome P450 homologues typically contained 100-200 ng of Columbia genomic DNA, 10 mM Tris-HCl (pH8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% (w/v) gelatin, 0.2 mM each dNTP, 312 ng "WAIGRDP" and 484 ng "Pet Haem" and 1.25 units Taq polymerase (Cetus). Reaction mixes (50 μL) were cycled 40 times between 95°C for 50 seconds, 45°C for 50 seconds and 72°C for 45 seconds.

The expected PCR product using the "WAIGRDP" and "Pet Haem" primers on a typical P450 gene template without an intron is around 150 base pairs. PCR fragments of approximately 140 to 155 base pairs were isolated and purified using the Mermaid® kit (BIO 101). The PCR fragments were re-amplified to obtain enough product for cloning and then end-repaired using Pfu DNA polymerase and finally cloned into pCR-Script™ Direct SK(+) (Stratagene). The ligated DNA was then transformed into competent DH5α cells (Inoue *et al.*, 1990).

Sequence of PCR products

Plasmid DNA from 15 transformants was prepared (Del Sal *et al.*, 1989). Sequencing data generated from these PCR fragments indicated that 11 out of the 15 represented unique clones. A distinct set of cytochrome P450 consensus amino acids was also found in the translated sequence within the *A. thaliana* PCR inserts, notably the

FXPeRF1 sequence The sequences of the PCR fragments were also compared with those of the petunia OGR-38 F3'H cDNA clone and the snapdragon sdF3'H cDNA clone. The PCR fragment, p58092.13, was most similar to the F3'H sequences from both petunia and snapdragon.

EXAMPLE 20-Screening of *A. thaliana* cDNA library

To isolate a cDNA clone of the p58092.13 PCR product, an *A. thaliana* cv. Columbia cDNA library (Newman *et al.*, 1994; D' Alessio *et al.*, 1992) was screened with a ^{32}P-labelled fragment of p58092.13 together with a ^{32}P-labelled fragment of the petunia *Ht1* cDNA insert (OGR-38) contained in pCGP1805 (Figure 8).

A total of 600,000 pfu was plated at a density of 50,000 plaques per 15 cm diameter plate as described by D' Alessio *et al* (1992). After incubation at 4°C overnight, duplicate lifts were taken onto Colony/Plaque Screen™ filters (DuPont) and treated as recommended by the manufacturer.

Prior to hybridization, the duplicate plaque lifts were washed in prewashing solution (50 mM Tris-HCl pH7.5, 1 M NaCl, 1 mM EDTA, 0.1% (w/v) sarcosine) at 65°C for 30 minutes; stripped in 0.4 M sodium hydroxide at 65°C for 30 minutes; then washed in a solution of 0.2 M Tris-HCl pH 8.0, 0.1 x SSC, 0.1% (w/v) SDS at 65°C for 30 minutes and finally rinsed in 2 x SSC, 1.0% (w/v) SDS.

Hybridization conditions included a prehybridization step in 50% (v/v) formamide, 1 M NaCl, 10% (w/v) dextran sulphate, 1% (w/v) SDS at 42°C for at least 1 hour. The ^{32}P-labelled fragment of p58092.13 ($2x10^6$cpm/mL) was then added to the hybridization solution and hybridization was continued at 42°C for a further 16 hours. The filters were then washed in 2 x SSC, 1% (w/v) SDS at 42°C for 2 x 1 hour and exposed to Kodak XAR film with an intensifying screen at -70°C for 16 hours.

Eleven strongly-hybridizing plaques were picked into PSB and rescreened as detailed above, to isolate purified plaques. These filters were also probed with ^{32}P-labelled fragment of the petunia *Ht1* cDNA insert (OGR-38) contained in pCGP1805 under low

stringency conditions. Low stringency conditions included prehybridization and hybridization at 42°C in 20% (v/v) formamide, 1 M NaCl, 10% (w/v) dextran sulphate, 1% (w/v) SDS and washing in 6xSSC, 1% (w/v) SDS (w/v) at 65°C for 1 hour.

The OGR-38 and p58092.13 probes hybridized with identical plaques. The 11 pure plaques were picked into PSB and the plasmid vectors pZL1 containing the cDNA clones were rescued using the bacterial strain DH10B(Zip). Plasmid DNA was prepared (Del Sal *et al.*, 1989) and the cDNA inserts were released upon digestion with *Bam*HI and *EcoR*I. The 11 plasmids contained cDNA inserts of between 800bp and 1 kb. Sequence data generated from the 5′ region of the cDNA inserts suggested that nine of these clones were identical. Sequence data was generated from the 5′ ends of the nine cDNA inserts and the 3′ end of only one cDNA insert. The sequence data generated from all clones was compiled to produce the nucleotide and translated sequence shown in Figures 22a and 22b.

The *A. thaliana* putative F3′H sequence was compared with the sequence of the petunia OGR-38 F3′H cDNA clone and is 64.7% similar to the petunia F3′H cDNA clone at the nucleic acid level, over 745 nucleotides and 63.7 % similar at the amino acid level, over 248 amino acids.

EXAMPLE 21 -Isolation of putative F3′H cDNA clone from *Rosa hybrida*

In order to isolate a homologue of the petunia F3′H cDNA clone, a *Rosa hybrida* cv Kardinal petal cDNA library was screened with ^{32}P-labelled fragments of the petunia *Ht1* cDNA clone (OGR-38) contained in pCGP1805 (Figure 8) and snapdragon F3′H cDNA clone (sdF3′H) contained in pCGP246 (Figure 18).

Construction of a petal cDNA library from Kardinal

Total RNA was prepared from the buds of *Rosa hybrida* cv. Kardinal stage 2. At this stage, the tightly closed buds were 1.5 cm high and approximately 0.9 cm wide with pale pink petals.

Frozen tissue (1-3 g) was ground in liquid nitrogen with a mortar and pestle, placed in 25 mL pre-chilled Buffer A [0.2 M boric acid, 10 mM EDTA (sodium salt) (pH 7.6)] and homogenized briefly. The extract was mixed on a rotary shaker until it reached room temperature and an equal volume of phenol/chloroform (1:1 v/v), equilibrated with Buffer A, was added. After mixing for a further 10 minutes, the RNA preparation was centrifuged at 10,000 x *g* for 10 minutes at 20°C. The upper aqueous phase was retained and the phenol interface re-extracted as above. The aqueous phases were pooled and adjusted to 0.1 M sodium acetate (pH 6.0), 2.5 volumes 95% ethanol were added and the mixture was stored at -20°C overnight.

The preparation was centrifuged at 10,000 x *g* for 10 minutes at 4°C, the pellet dissolved gently in 20 mL Buffer B [25 mM boric acid, 1.25 mM EDTA (sodium salt), 0.1 M NaCl (pH 7.6)] and 0.4 volumes 2-butoxyethanol (2BE) were added. This solution was incubated on ice for 30 minutes. It was then centrifuged at 10,000 x *g* for 10 minutes at 0°C and the supernatant carefully collected. After addition of 1.0 volume of 2BE and incubation on ice for a further 30 minutes, the mixture was again centrifuged at 10,000 x *g* for 10 minutes at 0°C. The resulting pellet was gently washed with Buffer A:2BE (1:1 v/v), then with 70% (v/v) ethanol, 0.1 M potassium acetate and finally with 95% ethanol. The pellet was air dried and dissolved in 1 mL diethyl pyrocarbonate (DEPC)-treated water. This was adjusted to 3 M lithium chloride, left on ice for 60 minutes and centrifuged at 10,000 x *g* for 10 minutes at 0°C. The resultant pellet was washed twice with 3 M LiCl and then with 70% ethanol, 0.1 M potassium acetate.

The RNA pellet was then dissolved in 400 μL DEPC-treated water and extracted with an equal volume phenol/chloroform. The RNA mix was centrifuged at 10,000 x *g* for 5 minutes at 20°C, the aqueous phase collected and made to 0.1 M sodium acetate, and a further 2.5 volumes of 95% ethanol were added. After 30 minutes incubation on ice, the mix was centrifuged at 13,000 rpm (5,000 x *g*) for 20 minutes at 20°C and the RNA pellet resuspended gently in 400 μL DEPC-treated water.

Poly $(A)^+$ RNA was selected from the total RNA by Oligotex dT-30 (Takara, Japan) following the manufacturer's protocol. The cDNA was synthesized according to the method in Brugliera *et al.* (1994) and used to construct a non-directional petal cDNA library in the *EcoR*I site of λZAPII (Stratagene). The total number of recombinants obtained was 3.5×10^5.

After transfecting XL1-Blue cells, the packaged cDNA mixture was plated at 50,000 pfu per 15 cm diameter plate. The plates were incubated at 37°C for 8 hours, and the phage were eluted into 100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl pH 8.0, 0.01% (w/v) gelatin (Phage Storage Buffer (PSB)) (Sambrook *et al.*, 1989). Chloroform was added and the phage stored at 4°C as an amplified library.

200,000 pfu of the amplified library were plated onto NZY plates (Sambrook *et al.*, 1989) at a density of 10,000 plaques per 15 cm plate after transfecting XL1-Blue MRF′ cells, and incubated at 37°C for 8 hours. After incubation at 4°C overnight, duplicate lifts (labelled as group A and group B) were taken onto Colony/Plaque Screen™ filters (DuPont) and treated as recommended by the manufacturer.

Screening of Kardinal cDNA library for F3′H homologues

Prior to hybridization, the duplicate plaque lifts were washed in prewashing solution (50 mM Tris-HCl pH7.5, 1 M NaCl, 1 mM EDTA, 0.1% (w/v) sarcosine) at 65°C for 30 minutes; stripped in 0.4 M sodium hydroxide at 65°C for 30 minutes; then washed in a solution of 0.2 M Tris-HCl pH 8.0, 0.1 x SSC, 0.1% (w/v) SDS at 65°C for 30 minutes and finally rinsed in 2 x SSC, 1.0% (w/v) SDS.

The group A filters of the duplicate lifts from the Kardinal cDNA library were screened with ^{32}P-labelled fragments of an *Nco*I fragment from pCGP1805 containing the petunia *Ht1* (OGR-38) cDNA clone (Figure 8), while the group B filters were screened with ^{32}P-labelled fragments of *EcoR*I/*Ssp*I fragment from pCGP246 containing the snapdragon F3′H clone (Figure 18).

Hybridization conditions included a prehybridization step in 10% (v/v) formamide, 1 M NaCl, 10% (w/v) dextran sulphate, 1% (w/v) SDS at 42°C for at least 1 hour. The ^{32}P-labelled fragments ($2x10^6$cpm/mL) was then added to the hybridization solution and hybridization was continued at 42°C for a further 16 hours. The filters were then washed at 42°C in 2 x SSC, 1% (w/v) SDS for 2 hours followed by 1 x SSC, 1% (w/v) SDS for 1 hour and finally in 0.2 x SSC/1% (w/v) SDS for 2 hours. The filters were exposed to Kodak XAR film with an intensifying screen at -70°C for 16 hours.

Four strongly-hybridizing plaques (R1, R2, R3, R4) were picked into PSB and rescreened to isolate plaques. The plasmids contained in the λZAP bacteriophage vector were rescued and digested with *EcoR*I to release the cDNA inserts. Clone R1 contained a 1.0 kb insert while clones R2, R3 and R4 contained inserts of around 1.3 kb each. Sequence data was generated from the 3′ and 5′ ends of the R4 cDNA insert. The partial nucleotide and putative amino acid sequence of R4 is shown in Figures 23a and 23b.

The rose R4 putative F3′H sequence was compared with that of the petunia OGR-38 F3′H sequence. At the nucleotide level, the R4 cDNA clone showed 63.2% and 62.1% similarity over 389 nucleotides at the 5′ end and 330 nucleotides at the 3′ end, respectively. At the amino acid level, the R4 clone showed 65.4 % and 73.9% similarity over 130 amino acids at the 5′ end and 69 amino acids at the 3′ end, respectively.

Those skilled in the art, however, will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations of any two or more of said steps or features.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. *J. Mol. Biol.* **215**: 403-410, 1990.

Ashikari, T., Kiuchi-Goto, N., Tanaka, Y., Shibano, Y., Amachi, T., and Yoshizumi, H. *Appl. Microbiol. Biotechnol.* **30**: 515-520, 1989.

Bethesda Research Laboratories. BRL pUC host: *E. coli* DH5α competent cells. *Bethesda Res. Lab. Focus.* **8(2)**: 9, 1986.

Brugliera, F., Holton, T.A., Stevenson, T.W., Farcy, E., Lu, C-Y and Cornish, E.C. *Plant J.* **5(1)**: 81-92, 1994.

Comai, L., Moran, P. and Maslyar, D., *Plant Mol. Biol.* **15**: 373-381, 1990.

D'Alessio et al., *Focus*, **14**: 76-79, 1992

Dellaporta, S.J., Wood, J. and Hick, J.B., *Plant Mol. Biol. Rep.* **1**: 19-21, 1983.

Del Sal, G., Manfioletti, G. and Schneider, C. *Biotechniques*, **7(5)**: 514-519, 1989.

Doodeman, M., Gerats, A.G.M., Schram, A.W., De Vlaming, P. and Bianchi, F., *Theor. Appl. Genet.* **67**: 357-366, 1984.

Dooner, H.K., Robbins, T.R. and Joregensen, R.A. *Ann. Rev. Genet.* **25**: 173-199, 1991.

Ebel, J. and Hahlbrock, K., In: T*he Flavonoids: Advances in Research Since 1980.* Harbourne, J.B. (ed.), Academic Press, New York, USA, 641-679, 1988.

Forkmann, G. *Plant Breeding* **106**: 1-26, 1991.

Frohman, M.A., Dush, M.K., Martin, G.R. *Proc. Natl. Acad. Sci. USA* **85**: 8998-9002, 1988.

Gamborg, O.L., Miller, R.A. and Ojima, K., *Exp. Cell Res.* **50**: 151-158, 1968.

Garfinkel, D.J. and Nester, E.W. *J. Bacteriol.* **144**:732-743, 1980.

Hahlbrock, K. and Grisebach, H., *Annu. Rev. Plant Physiol.* **30**: 105-130, 1979.

Hanahan, D., *J. Mol. Biol.* **166**: 557-580, 1983.

Holton, T.A., Brugliera, F. Lester, D.R., Tanaka, Y., Hyland, C.D., Menting, J.G.T., Lu, C., Farcy, E., Stevenson, T.W. and Cornish, E.C., *Nature* , **366**, 276-279, 1993.

Holton, T.A. and Cornish, E.C. *Plant Cell*, **7**: 1071-1083, 1995.

Inoue, H., Nojima, H. and Okayama, H. *Gene*, **96**: 23-28, 1990.

Ito, H., Fukuda, Y., Murata, K. and Kimura, A. *J. Bacteriol.* **153**: 163-168, 1983.

Kozak, M. *J. Cell. Biol.* **108**: 229-241, 1989.

Lazo, G.R., Pascal, A.S. and Ludwig, R.A., *Bio/technology*, **9**: 963-967, 1991.

Liang, P. and Pardee, A.B. *Science*, **257**: 967-971, 1992

Liang, P., Averboukh, L. and Pardee, A.B. *Nucl. Acids Res.* **21**: 3269-3275, 1993

Marchuk, D., Drumm, M., Saulino, A., Collins, F.S. *Nucl. Acids Res.* **19:** 1154, 1990

Markham, K. R., *Techniques of flavonoid identification*. London: Academic Press, 1982.

Martin, C. and Gerats, T. In: *The molecular biology of flowering*. (Jordan, B.R. ed), UK, CAB International, 219-255, 1993.

Mizutani, M., Ward, E., DiMaio, J., Ohta, D., Ryals, J. and Sato, R. *Biochem. Biophys. Res. Commun.* **190**: 875-880, 1993.

Murashige, T. and Skoog, F., *Physiol. Plant* , **15**: 73-97, 1962.

Newman et al., 1994

Pearson, W.R. and Lipman, D.J., *Proc. Natl. Acad. Sci. USA* **85**: 2444-2448, 1988.

Sambrook, J., Fritsch, E.F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual* (2nd edition). Cold Spring Harbor Laboratory Press, USA, 1989.

Schenk, R.U. and Hilderbrandt, A.C., *Can. J. Bot.* **50**: 199-204, 1972.

Schram, A.W., Jonsson, L.M.V. and Bennink, G.J.H., Biochemistry of flavonoid synthesis in *Petunia* hybrida. In: *Petunia* Sink, K.C. (ed.), Springer-Verlag, Berlin, Germany, pp 68-75, 1984.

Stafford, H.A., *Flavonoid Metabolism.* CRC Press, Inc. Boca Raton, Florida, USA, 1990.

Stotz, G. and Forkmann, G. *Z. Naturforsch* **37c**: 19-23, 1982.

Tabak, A.J.H., Meyer, H. and Bennink, G.J.H., *Planta* **139**, 67-71, 1978.

Tanaka, Y., Ashikari, T., Shibano, Y., Amachi, T., Yoshizumi, H. and Matsubara, H. *J. Biochem.* **103**: 954-961, 1988.

Turpen, T.H. and Griffith, O.M. *BioTechniques*, **4**: 11-15, 1986.

van Tunen A.J. and Mol J.N.M. In: *Plant Biotechnology* (Grierson, D. ed.) Glasgow: Blackie, **2**: 9-31, 1990.

Wiering, H. and de Vlaming, P., Inheritance and Biochemistry of Pigments. In: *Petunia* Sink, K.C. (ed.), Springer-Verlag, Berlin, Germany, pp 49-65, 1984.

Wallroth, M., Gerats, A.G.M., Rogers, S.G., Fraley, R.T. and Horsch, R.B., *Mol. Gen. Genet.* **202**: 6-15, 1986.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: FLORIGENE LIMITED

(ii) TITLE OF INVENTION: GENETIC SEQUENCES ENCODING FLAVONOID PATHWAY ENZYMES AND USES THEREFOR

(iii) NUMBER OF SEQUENCES: 10

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: DAVIES COLLISON CAVE
(B) STREET: 1 LITTLE COLLINS STREET
(C) CITY: MELBOURNE
(D) STATE: VICTORIA
(E) COUNTRY: AUSTRALIA
(F) ZIP: 3000

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: AU PROVISIONAL
(B) FILING DATE: 01-MAR-1996

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: HUGHES, DR E JOHN L
(C) REFERENCE/DOCKET NUMBER: EJH/EK

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: +61 3 9254 2777
(B) TELEFAX: +61 3 9254 2770
(C) TELEX: AA 31787

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1789 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GCAGGAATTG GTGAACCCCA TAGAAGTAAA ATACTCCTAT CTTTATTTCA TGGAAATCTT 60

AAGCCTAATT CTGTACACCG TCATTTTCTC ATTTCTTCTA CAATTCATTC TTAGATCATT 120

TTTCCGTAAA CGTTACCCTT TACCATTACC ACCAGGTCCA AAACCATGGC CAATTATAGG 180

AAACCTAGTC CATCTTGGAC CCAAACCACA TCAATCAACT GCAGCCATGG CTCAAACTTA 240

TGGACCACTC ATGTATCTTA AGATGGGGTT CGTAGACGTG GTGGTTGCAG CCTCGGCATC 300

GGTTGCAGCT CAGTTCTTGA AAACTCATGA TGCTAATTTC TCGAGCCGTC CACCAAATTC 360

TGGTGCAGAA CATATGGCTT ATAATTATCA GGATCTTGTT TTTGCACCTT ATGGACCTAG 420

ATGGCGTATG CTTAGGAAAA TTTGCTCAGT TCACCTTTTC TCTACCAAGG CTTTAGATGA 480

CTTCCGCCAT GTCCGCCAGG ATGAAGTGAA AACACTGACG CGCGCACTAG CAAGTGCAGG 540

CCAAAAGCCA GTCAAATTAG GTCAGTTATT GAACGTGTGC ACGACGAACG CACTCGCGCG 600

AGTAATGCTA GGTAAGCGAG TATTTGCCGA CGGAAGTGGC GATGTTGATC CACAAGCGGC 660

GGAGTTCAAG TCAATGGTGG TGGAAATGAT GGTAGTCGCC GGTGTTTTTA ACATTGGTGA 720

TTTTATTCCG CAACTTAATT GGTTAGATAT TCAAGGTGTA GCCGCTAAAA TGAAGAAGCT 780

CCACGCGCGT TTCGACGCGT TCTTGACTGA TATACTTGAA GAGCATAAGG GTAAAATTTT 840

TGGAGAAATG AAAGATTTGT TGAGTACTTT GATCTCTCTT AAAAATGATG ATGCGGATAA 900

TGATGGAGGG AAACTCACTG ATACAGAAAT TAAAGCATTA CTTTTGAACT TGTTTGTAGC 960

TGGAACAGAC ACATCTTCTA GTACAGTTGA ATGGGCCATT GCTGAGCTTA TTCGTAATCC 1020

AAAAATACTA GCCCAAGCCC AGCAAGAGAT CGACAAAGTC GTTGGAAGGG ACCGGCTAGT 1080

TGGCGAATTG GACCTAGCCC AATTGACATA CTTGGAAGCT ATAGTCAAGG AAACCTTTCG 1140

GCTTCATCCA TCAACCCCTC TTTCACTTCC TAGAATTGCA TCTGAGAGTT GTGAGATCAA 1200

TGGCTATTTC ATTCCAAAAG GCTCAACGCT TCTCCTTAAT GTTTGGGCCA TTGCTCGTGA 1260

TCCAAATGCA TGGGCTGATC CATTGGAGTT TAGGCCTGAA AGGTTTTTGC CAGGAGGTGA 1320

GAAGCCCAAA GTTGATGTCC GTGGGAATGA CTTTGAAGTC ATACCATTTG GAGCTGGACG 1380

```
TAGGATTTGT GCTGGAATGA ATTTGGGTAT ACGTATGGTC CAGTTGATGA TTGCAACTTT    1440

AATACATGCA TTTAACTGGG ATTTGGTCAG TGGACAATTG CCGGAGATGT TGAATATGGA    1500

AGAAGCATAT GGGCTGACCT TACAACGGGC TGATCCATTG GTTGTGCACC CAAGGCCTCG    1560

CTTAGAAGCC CAAGCGTACA TTGGGTGAGC AGCAACAGCC CATGGAGATA ACATGAGTGT    1620

TAAATGTATG AGTCTCCATA TCTTGTTTAG TTTGTTTATG CTTTGGATTT AGTAGTTTTT    1680

ATATTGATAG ATCAATGTTT GCATTGTCAG TAAGAATATC CGTTGCTTGT TTCATTAACT    1740

CCAGGTGGAC AATAAAAGAA GTAATTTGTA TGAAAAAAAA AAAAAAAAA                1789
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 512 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Glu Ile Leu Ser Leu Ile Leu Tyr Thr Val Ile Phe Ser Phe Leu
1               5                  10                  15

Leu Gln Phe Ile Leu Arg Ser Phe Phe Arg Lys Arg Tyr Pro Leu Pro
            20                  25                  30

Leu Pro Pro Gly Pro Lys Pro Trp Pro Ile Ile Gly Asn Leu Val His
        35                  40                  45

Leu Gly Pro Lys Pro His Gln Ser Thr Ala Ala Met Ala Gln Thr Tyr
    50                  55                  60

Gly Pro Leu Met Tyr Leu Lys Met Gly Phe Val Asp Val Val Val Ala
65                  70                  75                  80

Ala Ser Ala Ser Val Ala Ala Gln Phe Leu Lys Thr His Asp Ala Asn
                85                  90                  95

Phe Ser Ser Arg Pro Pro Asn Ser Gly Ala Glu His Met Ala Tyr Asn
            100                 105                 110

Tyr Gln Asp Leu Val Phe Ala Pro Tyr Gly Pro Arg Trp Arg Met Leu
        115                 120                 125

Arg Lys Ile Cys Ser Val His Leu Phe Ser Thr Lys Ala Leu Asp Asp
    130                 135                 140

Phe Arg His Val Arg Gln Asp Glu Val Lys Thr Leu Thr Arg Ala Leu
145                 150                 155                 160

Ala Ser Ala Gly Gln Lys Pro Val Lys Leu Gly Gln Leu Leu Asn Val
                165                 170                 175
```

```
Cys Thr Thr Asn Ala Leu Ala Arg Val Met Leu Gly Lys Arg Val Phe
            180                 185                 190

Ala Asp Gly Ser Gly Asp Val Asp Pro Gln Ala Ala Glu Phe Lys Ser
        195                 200                 205

Met Val Val Glu Met Met Val Val Ala Gly Val Phe Asn Ile Gly Asp
    210                 215                 220

Phe Ile Pro Gln Leu Asn Trp Leu Asp Ile Gln Gly Val Ala Ala Lys
225                 230                 235                 240

Met Lys Lys Leu His Ala Arg Phe Asp Ala Phe Leu Thr Asp Ile Leu
                245                 250                 255

Glu Glu His Lys Gly Lys Ile Phe Gly Glu Met Lys Asp Leu Leu Ser
            260                 265                 270

Thr Leu Ile Ser Leu Lys Asn Asp Asp Ala Asp Asn Asp Gly Gly Lys
        275                 280                 285

Leu Thr Asp Thr Glu Ile Lys Ala Leu Leu Leu Asn Leu Phe Val Ala
    290                 295                 300

Gly Thr Asp Thr Ser Ser Ser Thr Val Glu Trp Ala Ile Ala Glu Leu
305                 310                 315                 320

Ile Arg Asn Pro Lys Ile Leu Ala Gln Ala Gln Gln Glu Ile Asp Lys
                325                 330                 335

Val Val Gly Arg Asp Arg Leu Val Gly Glu Leu Asp Leu Ala Gln Leu
            340                 345                 350

Thr Tyr Leu Glu Ala Ile Val Lys Glu Thr Phe Arg Leu His Pro Ser
        355                 360                 365

Thr Pro Leu Ser Leu Pro Arg Ile Ala Ser Glu Ser Cys Glu Ile Asn
    370                 375                 380

Gly Tyr Phe Ile Pro Lys Gly Ser Thr Leu Leu Leu Asn Val Trp Ala
385                 390                 395                 400

Ile Ala Arg Asp Pro Asn Ala Trp Ala Asp Pro Leu Glu Phe Arg Pro
                405                 410                 415

Glu Arg Phe Leu Pro Gly Gly Glu Lys Pro Lys Val Asp Val Arg Gly
            420                 425                 430

Asn Asp Phe Glu Val Ile Pro Phe Gly Ala Gly Arg Arg Ile Cys Ala
        435                 440                 445

Gly Met Asn Leu Gly Ile Arg Met Val Gln Leu Met Ile Ala Thr Leu
    450                 455                 460

Ile His Ala Phe Asn Trp Asp Leu Val Ser Gly Gln Leu Pro Glu Met
465                 470                 475                 480

Leu Asn Met Glu Glu Ala Tyr Gly Leu Thr Leu Gln Arg Ala Asp Pro
                485                 490                 495

Leu Val Val His Pro Arg Pro Arg Leu Glu Ala Gln Ala Tyr Ile Gly
            500                 505                 510
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1711 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CGAATTCCCC CCCCCCCACA CCATTCAATG CCTAAGTCCT CCATTTGCCG GCCTAATAAC 60

TAAAAGCCCA CTCTTTCCGA CCATCTATAC ATGCAACACC AATATTATTC TTTAATTACG 120

ATGGATGATA TTAGCATAAC CAGCTTATTG GTGCCATGTA CTTTTATATT AGGGTTCTTG 180

CTTCTATATT CCTTCCTCAA CAAAAAAGTA AAGCCACTGC CACCTGGACC GAAGCCATGG 240

CCCATCGTCG GAAATCTGCC ACATCTTGGG CCGAAGCCCC ACCAGTCGAT GGCGGCGCTG 300

GCACGGGTGC ACGGCCCATT AATTCATCTG AAGATGGGCT TTGTGCATGT GGTTGTGGCC 360

TCCTCAGCAT CCGTTGCGGA GAAATTTCTG AAGGTGCATG ACGCAAACTT CTCGAGCAGG 420

CCTCCCAATT CGGGTGCAAA ACACGTGGCC TACAACTATC AGGACTTGGT CTTTGCTCCT 480

TATGGCCCAC GCTGGCGGAT GCTCAGGAAA ATCTGTGCAC TCCACCTCTT CTCCGCCAAA 540

GCCTTGAACG ACTTCACACA CGTCAGACAG GATGAGGTGG GGATCCTCAC TCGCGTTCTA 600

GCAGATGCAG GAGAAACGCC GTTGAAATTA GGGCAGATGA TGAACACATG CGCCACCAAT 660

GCAATAGCGC GTGTTATGTT GGGTCGACGC GTGGTTGGAC ACGCAGACTC AAAGGCGGAG 720

GAGTTTAAGG CAATGGTAGT GGAGTTGATG GTATTAGCTG GTGTGTTCAA CTTAGGTGAT 780

TTTATCCCAC CTCTTGAAAA ATTGGATCTT CAAGGTGTCA TTGCTAAGAT GAAGAAGCTT 840

CACTTGCGTT TCGACTCGTT CTTGAGTAAG ATCCTTGGAG ACCACAAGAT CAACAGCTCA 900

GATGAAACCA AAGGCCATTC GGATTTGTTG AACATGTTAA TTTCTTTGAA GGACGCTGAT 960

GATGCCGAAG GAGGGAGGCT CACCGACGTA GAAATTAAAG CGTTGCTCTT GAACTTGTTT 1020

GCTGCAGGAA CTGACACAAC ATCAAGCACT GTGGAATGGT GCATAGCTGA GTTAGTACGA 1080

CATCCTGAAA TCCTTGCCCA AGTCCAAAAA GAACTCGACT CTGTTGTTGG TAAGAATCGG 1140

GTGGTGAAGG AGGCTGATCT GGCCGGATTA CCATTCCTCC AAGCGGTCGT CAAGGAAAAT 1200

TTCCGACTCC ATCCCTCCAC CCCGCTCTCC CTACCGAGGA TCGCACATGA GAGTTGTGAA 1260

GTGAATGGAT ACTTGATTCC AAAGGGTTCG ACACTTCTTG TCAATGTTTG GGCAATTGCT 1320

CGCGATCCAA ATGTGTGGGA TGAACCACTA GAGTTCCGGC CTGAACGATT CTTGAAGGGC 1380

GGGGAAAAGC CTAATGTCGA TGTTAGAGGG AATGATTTCG AATTGATACC GTTCGGAGCG 1440

GGCCGAAGAA TTTGTGCAGG AATGAGCTTA GGAATACGTA TGGTCCAGTT GTTGACAGCA 1500

ACTTTGAACC ATGCGTTTGA CTTTGATTTG GCGGATGGAC AGTTGCCTGA AAGCTTAAAC 1560

ATGGAGGAAG CTTATGGGCT GACCTTGCAA CGAGCTGACC CTTTGGTAGT GCACCCGAAG 1620

CCTAGGTAGG CACCTCATGT TTATCAAACT TACGACTCAT GTTTAGAGAA CCTCTTGTTG 1680

TTTTATCAGA TTGAAGTGTG ATGTCCAAGA C 1711

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 512 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Gln His Gln Tyr Tyr Ser Leu Ile Thr Met Asp Asp Ile Ser Ile
1 5 10 15

Thr Ser Leu Leu Val Pro Cys Thr Phe Ile Leu Gly Phe Leu Leu Leu
20 25 30

Tyr Ser Phe Leu Asn Lys Lys Val Lys Pro Leu Pro Pro Gly Pro Lys
35 40 45

Pro Trp Pro Ile Val Gly Asn Leu Pro His Leu Gly Pro Lys Pro His
50 55 60

Gln Ser Met Ala Ala Leu Ala Arg Val His Gly Pro Leu Ile His Leu
65 70 75 80

Lys Met Gly Phe Val His Val Val Val Ala Ser Ser Ala Ser Val Ala
85 90 95

Glu Lys Phe Leu Lys Val His Asp Ala Asn Phe Ser Ser Arg Pro Pro
100 105 110

Asn Ser Gly Ala Lys His Val Ala Tyr Asn Tyr Gln Asp Leu Val Phe
115 120 125

Ala Pro Tyr Gly Pro Arg Trp Arg Met Leu Arg Lys Ile Cys Ala Leu
130 135 140

His Leu Phe Ser Ala Lys Ala Leu Asn Asp Phe Thr His Val Arg Gln
145 150 155 160

Asp Glu Val Gly Ile Leu Thr Arg Val Leu Ala Asp Ala Gly Glu Thr
165 170 175

Pro Leu Lys Leu Gly Gln Met Met Asn Thr Cys Ala Thr Asn Ala Ile
180 185 190

```
Ala Arg Val Met Leu Gly Arg Arg Val Val Gly His Ala Asp Ser Lys
        195                 200                 205

Ala Glu Glu Phe Lys Ala Met Val Val Glu Leu Met Val Leu Ala Gly
    210                 215                 220

Val Phe Asn Leu Gly Asp Phe Ile Pro Pro Leu Glu Lys Leu Asp Leu
225                 230                 235                 240

Gln Gly Val Ile Ala Lys Met Lys Lys Leu His Leu Arg Phe Asp Ser
                245                 250                 255

Phe Leu Ser Lys Ile Leu Gly Asp His Lys Ile Asn Ser Ser Asp Glu
            260                 265                 270

Thr Lys Gly His Ser Asp Leu Leu Asn Met Leu Ile Ser Leu Lys Asp
        275                 280                 285

Ala Asp Asp Ala Glu Gly Gly Arg Leu Thr Asp Val Glu Ile Lys Ala
    290                 295                 300

Leu Leu Leu Asn Leu Phe Ala Ala Gly Thr Asp Thr Thr Ser Ser Thr
305                 310                 315                 320

Val Glu Trp Cys Ile Ala Glu Leu Val Arg His Pro Glu Ile Leu Ala
                325                 330                 335

Gln Val Gln Lys Glu Leu Asp Ser Val Val Gly Lys Asn Arg Val Val
            340                 345                 350

Lys Glu Ala Asp Leu Ala Gly Leu Pro Phe Leu Gln Ala Val Val Lys
        355                 360                 365

Glu Asn Phe Arg Leu His Pro Ser Thr Pro Leu Ser Leu Pro Arg Ile
    370                 375                 380

Ala His Glu Ser Cys Glu Val Asn Gly Tyr Leu Ile Pro Lys Gly Ser
385                 390                 395                 400

Thr Leu Leu Val Asn Val Trp Ala Ile Ala Arg Asp Pro Asn Val Trp
                405                 410                 415

Asp Glu Pro Leu Glu Phe Arg Pro Glu Arg Phe Leu Lys Gly Gly Glu
            420                 425                 430

Lys Pro Asn Val Asp Val Arg Gly Asn Asp Phe Glu Leu Ile Pro Phe
        435                 440                 445

Gly Ala Gly Arg Arg Ile Cys Ala Gly Met Ser Leu Gly Ile Arg Met
    450                 455                 460

Val Gln Leu Leu Thr Ala Thr Leu Asn His Ala Phe Asp Phe Asp Leu
465                 470                 475                 480

Ala Asp Gly Gln Leu Pro Glu Ser Leu Asn Met Glu Glu Ala Tyr Gly
                485                 490                 495

Leu Thr Leu Gln Arg Ala Asp Pro Leu Val Val His Pro Lys Pro Arg
            500                 505                 510
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 971 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GATATGCTTA GCACTTTAAT CTCCCTTAAA GGAACTGATC TTGACGGTGA CGGAGGAAGC 60

TTAACGGATA CTGAGATTAA AGCCTTGCTA TTGAACATGT TCACAGCTGG AACTGACACG 120

TCAGCAAGTA CGGTGGACTG GGCTATAGCT GAACTTATCC GTCACCCGGA TATAATGGTT 180

AAAGCCCAAG AAGAACTTGA TATTGTTGTG GGCCGTGACA GGCCTGTTAA TGAATCAGAC 240

ATCGCTCAGC TTCCTTACCT TCAGGCGGTT ATCAAAGAGA ATTTCAGGCT TCATCCACCA 300

ACACCACTCT CGTTACCACA CATCGCGTCA GAGAGCTGTG AGATCAACGG CTACCATATC 360

CCGAAAGGAT CGACTCTATT TGACGGACAT ATGGGCCTAG GCCGTGACCC GGATCAATGG 420

TCCGACCCGT TAGCATTTAA ACCCGAGAGA TTCTTACCCG GTGGTGAAAA ATCCGGCGTT 480

GATGTGAAAG GAAGCGATTT CGAGCTAATA CCGTTCGGGG CTGGGAGGCC AATCTGTGCA 540

GGTTTAAGTT TAGGGCTACG TACAGATTTA AGTTGCCTTC ACGCCAACGT TGCTCACAAG 600

CATTTGATTG GGAACTTCAG CTGGAGAAGT TACGCCGGAC AACCTGAATA TCGCAGGAAA 660

AGTTTACTGG GCTTTACACT GCAAAGAGCG GTTCCTTCGG TGGTACACCC TAAGCCAAGG 720

TTGGCCCCGA ACGTTTATGG ACCCCGGGTC GGCTTAAAAT TTAACTTTGC TTCTTGGACA 780

AGGTATATGG CTTGCACGAA ACTAACGTTT TAACACACCG TAGTTTGATC CGGAGTTAGC 840

TTTATGTAAG AACGTGTAAC GCCAAATCAA GCCATTATCA ACTACCGTGA GCTGTTTGTA 900

CCCTATCTAT AAATCTTGAA GAGGAACATT TCAGAACTCT TGACTATGTT TCAGGAACAA 960

AAAAAAAAAA A 971

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 270 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Asp Met Leu Ser Thr Leu Ile Ser Leu Lys Gly Thr Asp Leu Asp Gly
1               5                  10                  15

Asp Gly Gly Ser Leu Thr Asp Thr Glu Ile Lys Ala Leu Leu Leu Asn
            20                  25                  30

Met Phe Thr Ala Gly Thr Asp Thr Ser Ala Ser Thr Val Asp Trp Ala
        35                  40                  45

Ile Ala Glu Leu Ile Arg His Pro Asp Ile Met Val Lys Ala Gln Glu
    50                  55                  60

Glu Leu Asp Ile Val Val Gly Arg Asp Arg Pro Val Asn Glu Ser Asp
65                  70                  75                  80

Ile Ala Gln Leu Pro Tyr Leu Gln Ala Val Ile Lys Glu Asn Phe Arg
                85                  90                  95

Leu His Pro Pro Thr Pro Leu Ser Leu Pro His Ile Ala Ser Glu Ser
            100                 105                 110

Cys Glu Ile Asn Gly Tyr His Ile Pro Lys Gly Ser Thr Leu Phe Asp
        115                 120                 125

Gly His Met Gly Leu Gly Arg Asp Pro Asp Gln Trp Ser Asp Pro Leu
    130                 135                 140

Ala Phe Lys Pro Glu Arg Phe Leu Pro Gly Gly Glu Lys Ser Gly Val
145                 150                 155                 160

Asp Val Lys Gly Ser Asp Phe Glu Leu Ile Pro Phe Gly Ala Gly Arg
                165                 170                 175

Pro Ile Cys Ala Gly Leu Ser Leu Gly Leu Arg Thr Asp Leu Ser Cys
            180                 185                 190

Leu His Ala Asn Val Ala His Lys His Leu Ile Gly Asn Phe Ser Trp
        195                 200                 205

Arg Ser Tyr Ala Gly Gln Pro Glu Tyr Arg Arg Lys Ser Leu Leu Gly
    210                 215                 220

Phe Thr Leu Gln Arg Ala Val Pro Ser Val Val His Pro Lys Pro Arg
225                 230                 235                 240

Leu Ala Pro Asn Val Tyr Gly Pro Arg Val Gly Leu Lys Phe Asn Phe
                245                 250                 255
```

Ala Ser Trp Thr Arg Tyr Met Ala Cys Thr Lys Leu Thr Phe
260 265 270

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 458 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CACATTCCCA AAAACACCAC TTTATTGGTA AATGTGTGGG CCATCGCACG CGACCCTGAG 60

GTTTGGGCCG ACCCGTTAGA GTTTAAACCC GAAAGATTTT TGCCGGGCGG CGAAAAGCCC 120

AATGTGGATG TGAAAGGAAA CGATTTTGAG CTGATTCCGT TCGGGGCGGG CCGACGGATT 180

TGTGCTGGGC TGAGTTTGGG CCTGCGTATG GTCCAGTTGA TGACAGCCAC TTTGGCCCAT 240

ACTTATGATT GGGCCTTAGC TGATGGGCTT ATGCCCGAAA AGCTTAACAT GGATGAGGCT 300

TATGGGCTTA CCTTACAGCG TAAGGTGCCA CTTAATGGTC CACCCGACCC CGTCGGCTTC 360

TCGGCCCGTG TTTAATAATT CCGGGGTTTT TAAAAGCGGG TTACTTTTGT TTATGTATTA 420

TTCCGTACTA GTTTGAAAAT AATGGTATTA GAGAAATG 458

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 124 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

His Ile Pro Lys Asn Thr Thr Leu Leu Val Asn Val Trp Ala Ile Ala
1 5 10 15

Arg Asp Pro Glu Val Trp Ala Asp Pro Leu Glu Phe Lys Pro Glu Arg
20 25 30

Phe Leu Pro Gly Gly Glu Lys Pro Asn Val Asp Val Lys Gly Asn Asp
35 40 45

Phe Glu Leu Ile Pro Phe Gly Ala Gly Arg Arg Ile Cys Ala Gly Leu
50 55 60

```
Ser Leu Gly Leu Arg Met Val Gln Leu Met Thr Ala Thr Leu Ala His
65                  70                  75                  80

Thr Tyr Asp Trp Ala Leu Ala Asp Gly Leu Met Pro Glu Lys Leu Asn
                85                  90                  95

Met Asp Glu Ala Tyr Gly Leu Thr Leu Gln Arg Lys Val Pro Leu Asn
            100                 105                 110

Gly Pro Pro Asp Pro Val Gly Phe Ser Ala Arg Val
        115                 120
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 791 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
AGCGCATGCC TTGGCAAATT CAGGGTCAAA GGTAGTGAAC CTGGCGCAAC TGCTGAACCT      60

GTGCACGGTC AATGCTCTAG GAAGGGTGAT GGTAGGGCGG AGGGTTTTCG GCGACGGCAG     120

CGGAGGCGAC GATCCGAAGG CGGACGAGTT CAAATCGATG GTGGTGGAGA TGATGGTGTT     180

GGCAGGAGTG TTCAACATAG GTGACTTCAT CCCCTCTCTC GAATGGCTTG ACTTGCAAGG     240

CGTGGCGTCC AAGATGAAGA AGCTCCACAA GAGATTCGAC GACTTCTTGA CAGCCATTGT     300

CGAGGACCAC AAGAAGGGCT CCGGCACGGC GGGGCACGTC GACATGTTGA CCACTCTGCT     360

CTCGCTCAAG GAAGACGCCG ACGGCAAGGA GTGCCGGGCG AAGAATATGT GCCGGGATGA     420

GCTTGGGCCT CCGTATGGTC CATTTAATGA CTGCAACATT GGTCCACGCA TTTAATTGGG     480

CCTTGGCTGA TGGGCTGACC GCTGAGAAGT TAAACATGGA TGAAGCATAT GGGCTCACTC     540

TACAACGAGC TGCACCGTTA ATGGTGCACC CGCGCACCAG GCTGGCCCCA CAGGCATATA     600

AAACTTCATC ATCTTAATTA GAGAGCTATG TTCTGGGTGT GCCCGGTTTG ATGTCTCCAT     660

GTTTTCTATT TAGGTTTAAA TCTGTAAGAT AAGGTGATTC TATGCTGAAT CACAAAAGTT     720

GCTATCTAAA TTCCATGTCC AATGAAACCG TTCTTCTTCC CTTCTTATAA TTTATGAATA     780

CTTATGAAAA A                                                          791
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 204 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Ala His Ala Leu Ala Asn Ser Gly Ser Lys Val Val Asn Leu Ala Gln
1               5                  10                  15

Leu Leu Asn Leu Cys Thr Val Asn Ala Leu Gly Arg Val Met Val Gly
            20                  25                  30

Arg Arg Val Phe Gly Asp Gly Ser Gly Gly Asp Asp Pro Lys Ala Asp
        35                  40                  45

Glu Phe Lys Ser Met Val Val Glu Met Met Val Leu Ala Gly Val Phe
    50                  55                  60

Asn Ile Gly Asp Phe Ile Pro Ser Leu Glu Trp Leu Asp Leu Gln Gly
65                  70                  75                  80

Val Ala Ser Lys Met Lys Lys Leu His Lys Arg Phe Asp Asp Phe Leu
                85                  90                  95

Thr Ala Ile Val Glu Asp His Lys Lys Gly Ser Gly Thr Ala Gly His
            100                 105                 110

Val Asp Met Leu Thr Thr Leu Leu Ser Leu Lys Glu Asp Ala Asp Gly
        115                 120                 125

Lys Glu Ala Gly Arg Arg Ile Cys Ala Gly Met Ser Leu Gly Leu Arg
    130                 135                 140

Met Val His Leu Met Thr Ala Thr Leu Val His Ala Phe Asn Trp Ala
145                 150                 155                 160

Leu Ala Asp Gly Leu Thr Ala Glu Lys Leu Asn Met Asp Glu Ala Tyr
                165                 170                 175

Gly Leu Thr Leu Gln Arg Ala Ala Pro Leu Met Val His Pro Arg Thr
            180                 185                 190

Arg Leu Ala Pro Gln Ala Tyr Lys Thr Ser Ser Ser
        195                 200
```


DATED this 1st day of March, 1996
FLORIGENE LIMITED
By Its Patent Attorneys
DAVIES COLLISON CAVE

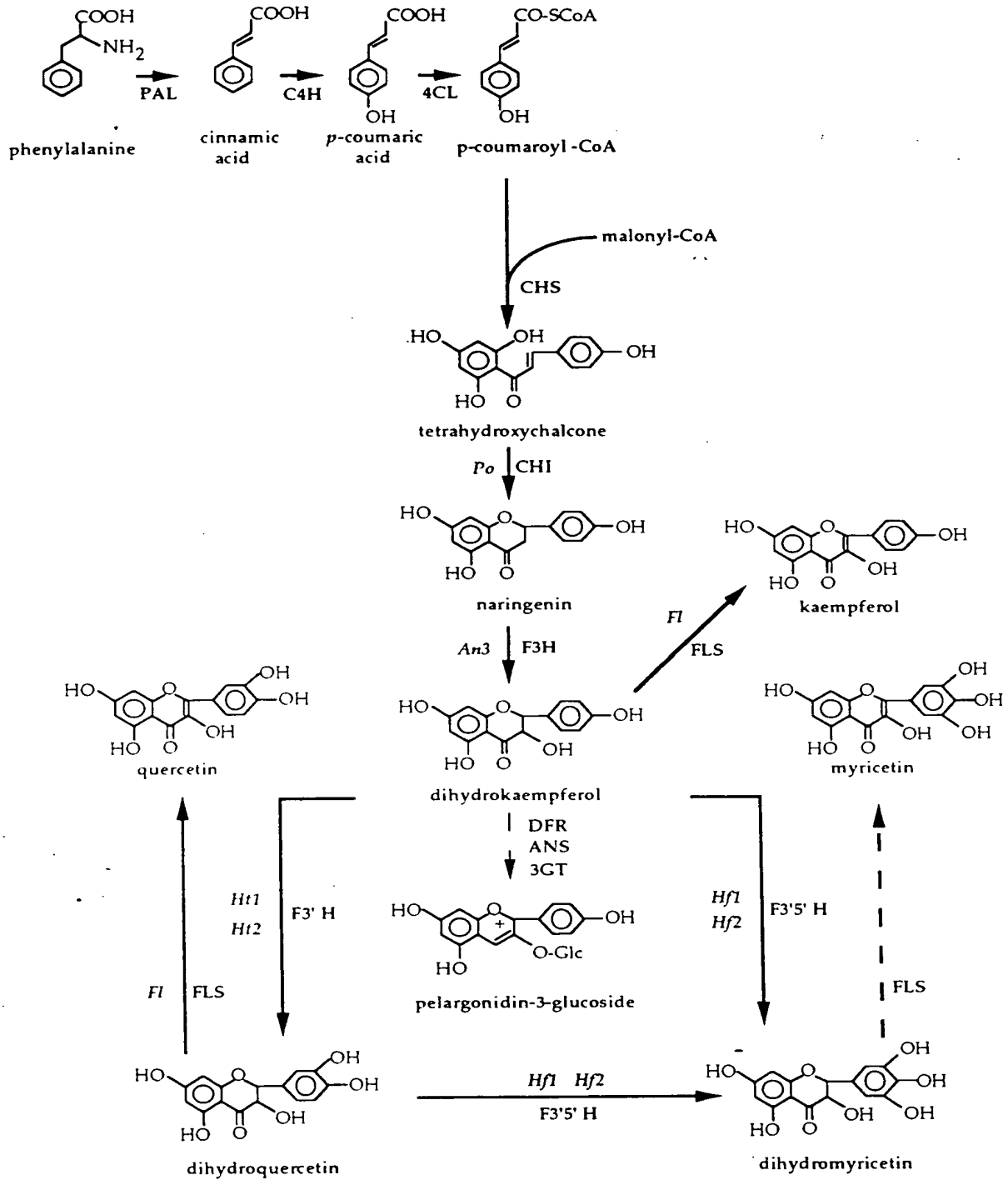

COOH
NH2
PAL
phenylalanine
COOH
C4H
cinnamic acid
COOH
OH
4CL
p-coumaric acid
CO-SCoA
OH
p-coumaroyl -CoA
malonyl-CoA
CHS
tetrahydroxychalcone
Po
CHI
naringenin
An3
F3H
dihydrokaempferol
Fl
FLS
kaempferol
quercetin
myricetin
DFR
ANS
3GT
O-Glc
pelargonidin-3-glucoside
Ht1
Ht2
F3' H
Fl
FLS
Hf1
Hf2
F3'5' H
FLS
dihydroquercetin
dihydromyricetin

Figure 1a

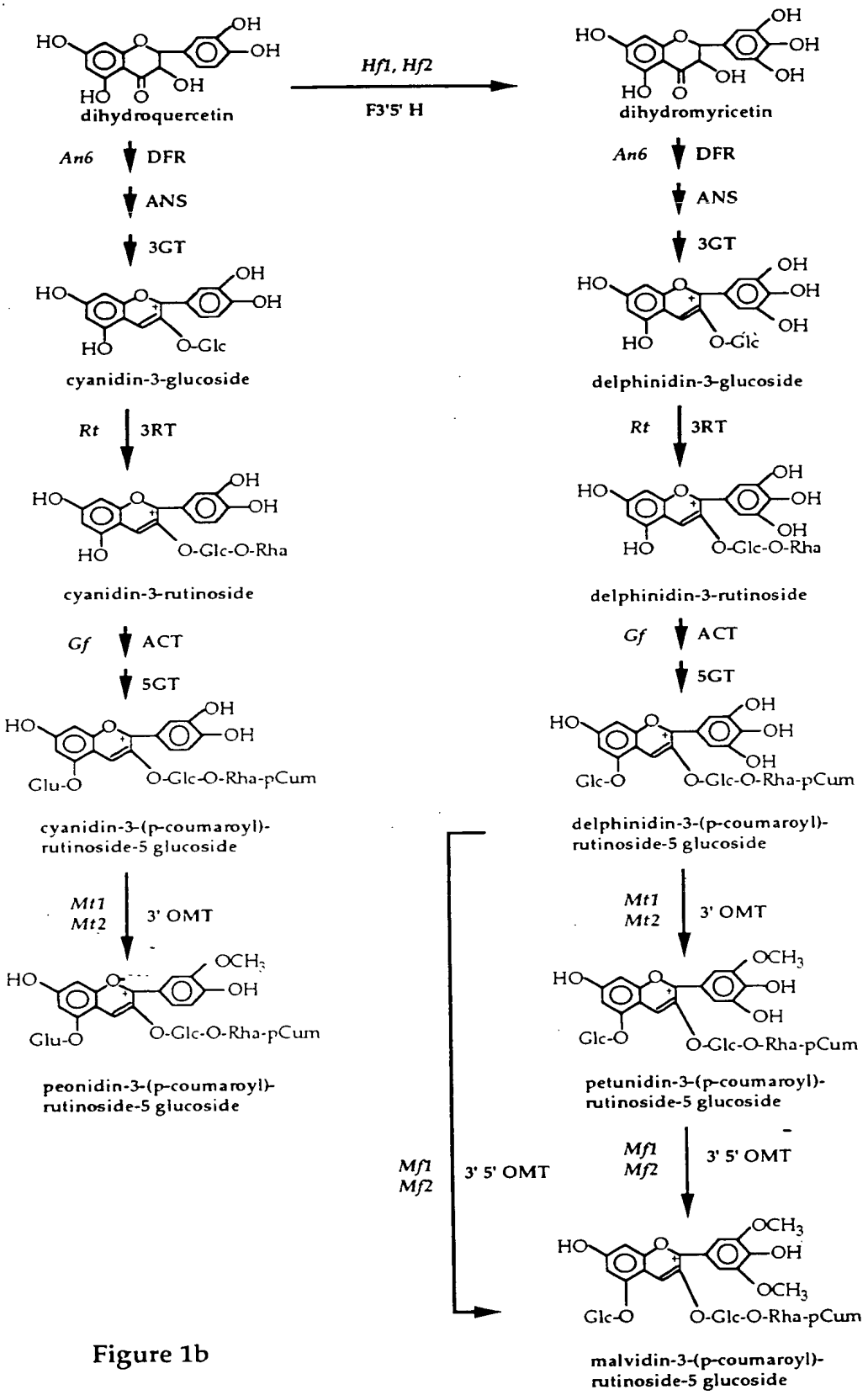

dihydroquercetin
Hf1, Hf2
F3'5' H
dihydromyricetin
An6
DFR
ANS
3GT
cyanidin-3-glucoside
delphinidin-3-glucoside
Rt
3RT
cyanidin-3-rutinoside
delphinidin-3-rutinoside
Gf
ACT
5GT
cyanidin-3-(p-coumaroyl)-rutinoside-5 glucoside
delphinidin-3-(p-coumaroyl)-rutinoside-5 glucoside
Mt1
Mt2
3' OMT
peonidin-3-(p-coumaroyl)-rutinoside-5 glucoside
petunidin-3-(p-coumaroyl)-rutinoside-5 glucoside
Mf1
Mf2
3' 5' OMT
malvidin-3-(p-coumaroyl)-rutinoside-5 glucoside

Figure 1b

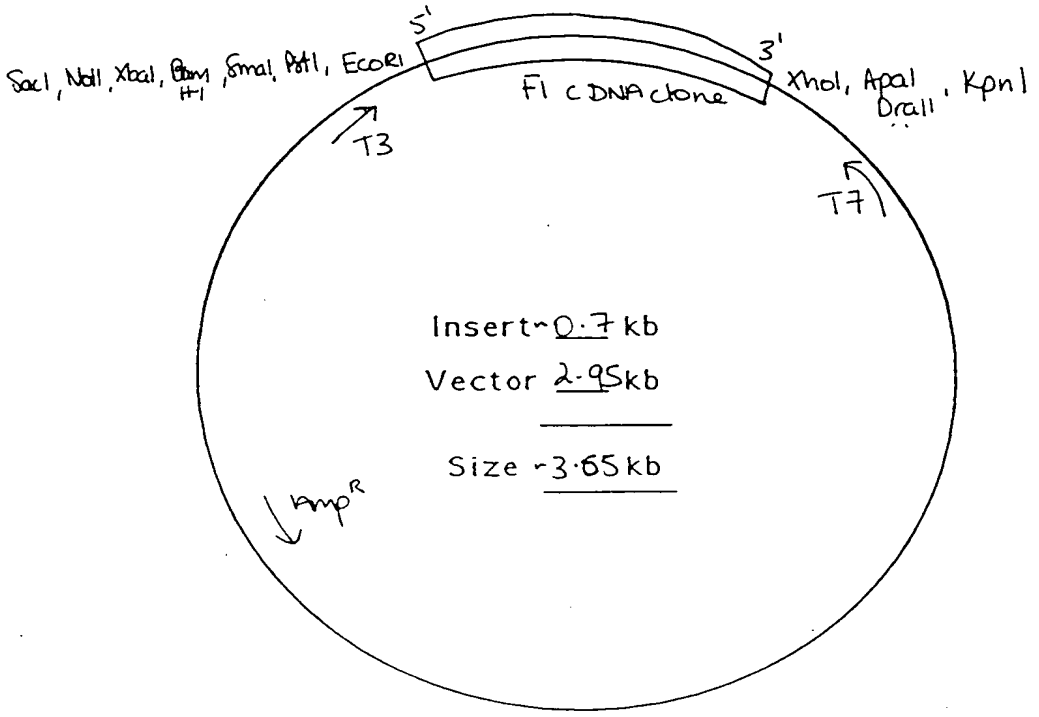
PLASMID pCGP 161

Constructed by _____ Date _____

Purpose C4H cDNA clone from P. hybrida

Phenotype Amp^R

MAP



Replicon/Site pBluescript SK II

Insert source cDNA prepared from 06B petal RNA (stages 3-4)

Lab Book pp CH BK 4 p 20247

Comments bp sequenced from 5' end

83.1% sequence similarity with mungbean C4H at nucleotide level (over 295 nucleotides) and 93.9% at amino acid level.

Figure 2

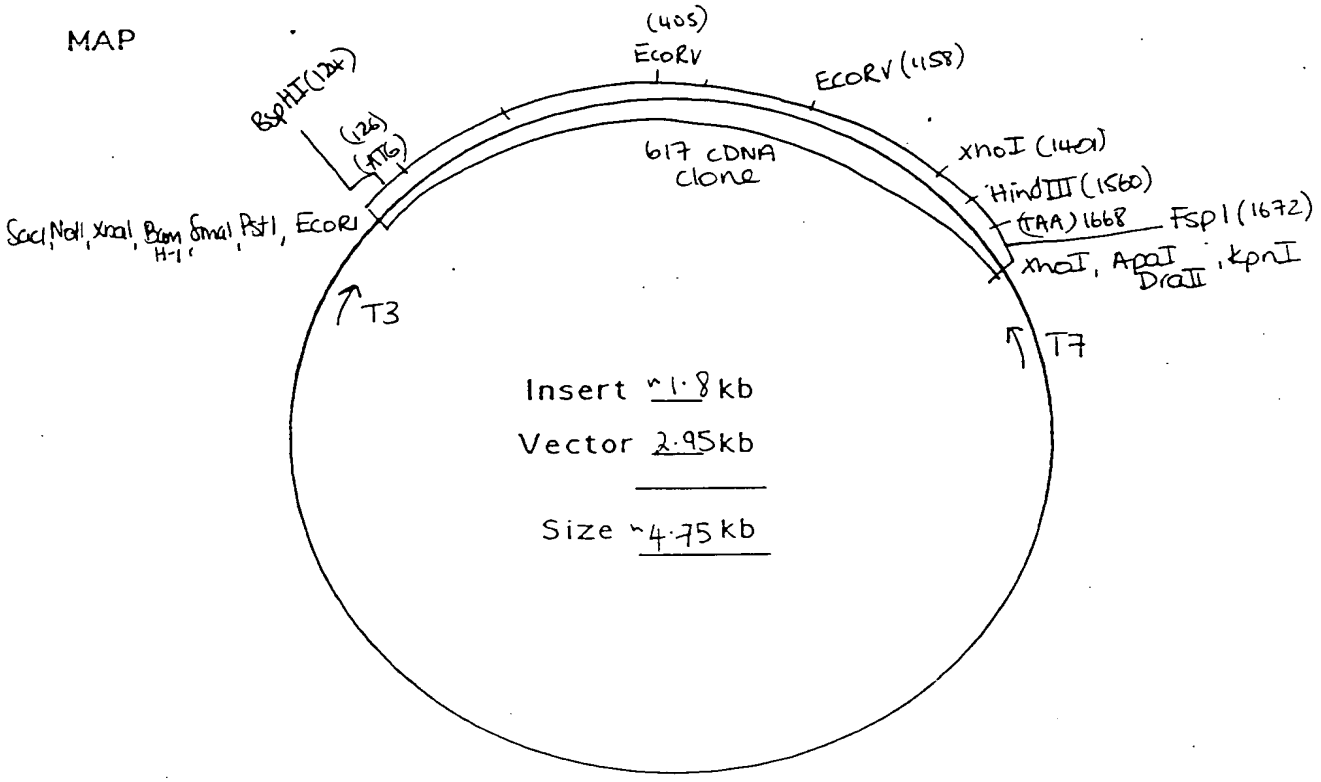
PLASMID pCGP602

Constructed by Yoshi Tanaka Date _____

Purpose Hf1 cDNA clone

Phenotype Amp^R

MAP



Replicon/Site pBluescript SK II / λZAP II

Insert source cDNA prepared from OGB petal RNA (Stages 3-4)

Lab Book pp 20951 (Yoshi)

Comments Full-length Hf1 cDNA (Clone #617)

Figure 3

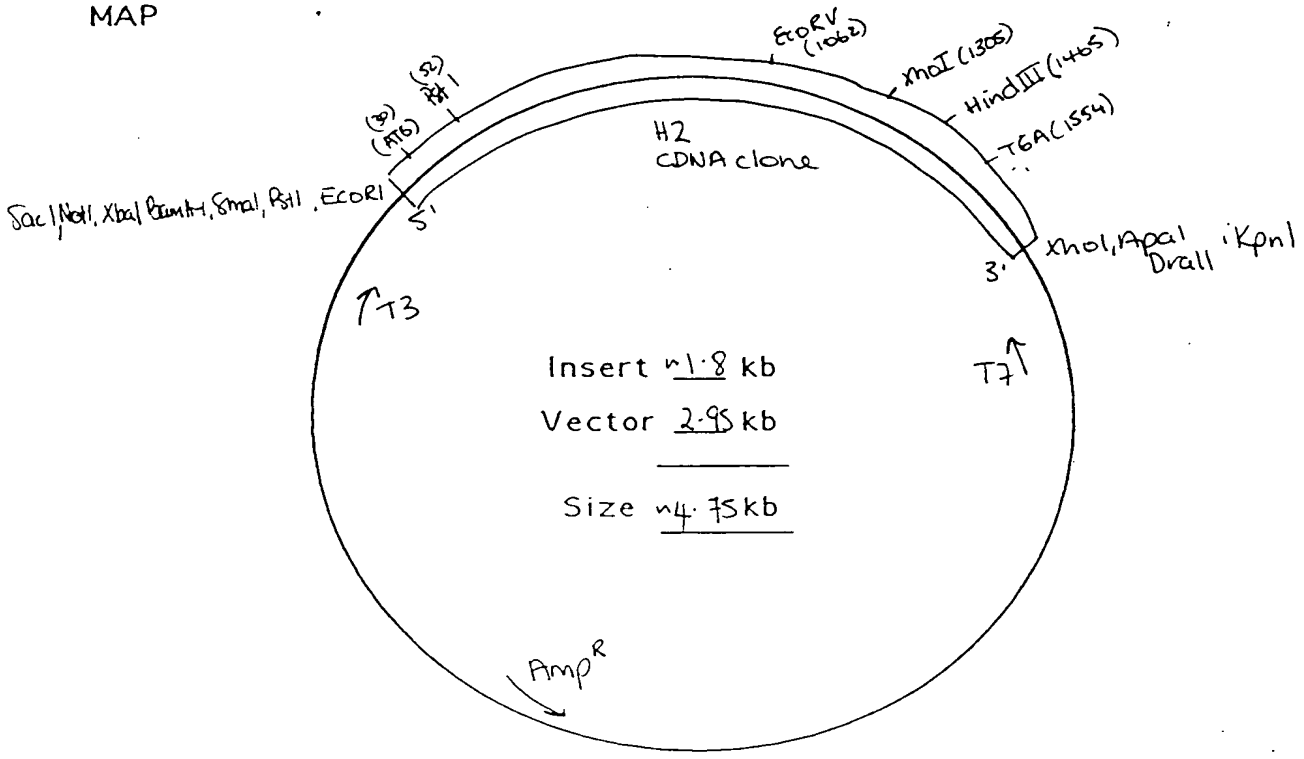
PLASMID pCGP175

Constructed by _____ Date _____

Purpose ? Hf2 cDNA clone (H2)

Phenotype Amp^R

MAP



Replicon/Site pBluescript SK

Insert source cDNA prepared from OGB petal RNA (Stages 3-4)

Lab Book pp _____

Comments H2 cDNA clone corresponding to Hf2 locus

Figure 4

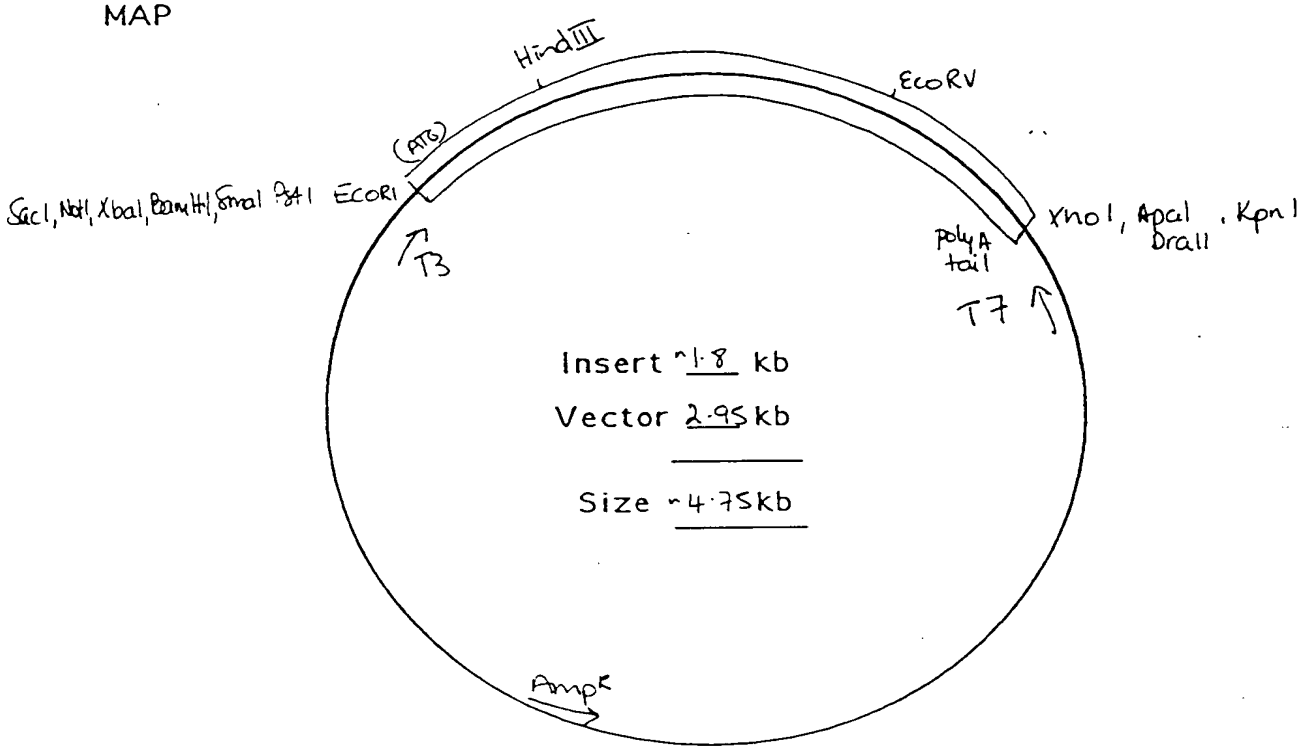
PLASMID pCGP 619

Constructed by Yoshi Tanaka Date _____

Purpose 651 cDNA clone = cytochrome P450

Phenotype Amp^R

MAP



Replicon/Site pBluescript SK

Insert source cDNA prepared from 06B petal RNA (Stages 3-4)

Lab Book pp 22252

Comments 651 cDNA clone

not linked to Ht1 - has F3'H activity in yeast.

Figure 5

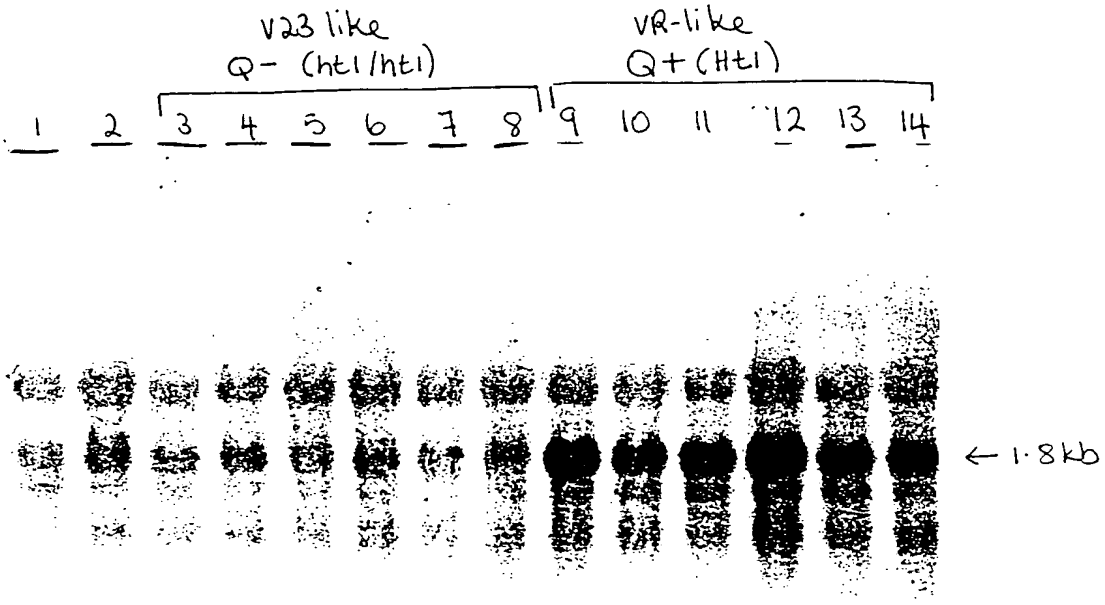


Figure 6

OGR-38 Sequence Range: 1 to 1789

```
                10          20          30          40
                *   *   *   *   *   *   *   *   *   *
g cag gaa ttg gtg aac ccc ata gaa gta aaa tac tcc tat ctt tat ttc

50          60          70          80          90
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
ATG GAA ATC TTA AGC CTA ATT CTG TAC ACC GTC ATT TTC TCA TTT CTT
 M   E   I   L   S   L   I   L   Y   T   V   I   F   S   F   L>

100         110         120         130         140
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
CTA CAA TTC ATT CTT AGA TCA TTT TTC CGT AAA CGT TAC CCT TTA CCA
 L   Q   F   I   L   R   S   F   F   R   K   R   Y   P   L   P>

   150         160         170         180         190
   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *
TTA CCA CCA GGT CCA AAA CCA TGG CCA ATT ATA GGA AAC CTA GTC CAT
 L   P   P   G   P   K   P   W   P   I   I   G   N   L   V   H>

      200         210         220         230         240
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
CTT GGA CCC AAA CCA CAT CAA TCA ACT GCA GCC ATG GCT CAA ACT TAT
 L   G   P   K   P   H   Q   S   T   A   A   M   A   Q   T   Y>

        250         260         270         280
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
GGA CCA CTC ATG TAT CTT AAG ATG GGG TTC GTA GAC GTG GTG GTT GCA
 G   P   L   M   Y   L   K   M   G   F   V   D   V   V   V   A>

290         300         310         320         330
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
GCC TCG GCA TCG GTT GCA GCT CAG TTC TTG AAA ACT CAT GAT GCT AAT
 A   S   A   S   V   A   A   Q   F   L   K   T   H   D   A   N>

340         350         360         370         380
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
TTC TCG AGC CGT CCA CCA AAT TCT GGT GCA GAA CAT ATG GCT TAT AAT
 F   S   S   R   P   P   N   S   G   A   E   H   M   A   Y   N>

    400         410         420         430
    *   *   *   *   *   *   *   *   *   *   *   *   *   *   *
TAT CAG GAT CTT GTT TTT GCA CCT TAT GGA CCT AGA TGG CGT ATG CTT
 Y   Q   D   L   V   F   A   P   Y   G   P   R   W   R   M   L>

    440         450         460         470         480
    *   *   *   *   *   *   *   *   *   *   *   *   *   *   *
AGG AAA ATT TGC TCA GTT CAC CTT TTC TCT ACC AAG GCT TTA GAT GAC
 R   K   I   C   S   V   H   L   F   S   T   K   A   L   D   D>

        490         500         510         520
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
TTC CGC CAT GTC CGC CAG GAT GAA GTG AAA ACA CTG ACG CGC GCA CTA
 F   R   H   V   R   Q   D   E   V   K   T   L   T   R   A   L>

530         540         550         560         570
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
GCA AGT GCA GGC CAA AAG CCA GTC AAA TTA GGT CAG TTA TTG AAC GTG
 A   S   A   G   Q   K   P   V   K   L   G   Q   L   L   N   V>

580         590         600         610         620
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
TGC ACG ACG AAC GCA CTC GCG CGA GTA ATG CTA GGT AAG CGA GTA TTT
 C   T   T   N   A   L   A   R   V   M   L   G   K   R   V   F>

   630         640         650         660         670
   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *
GCC GAC GGA AGT GGC GAT GTT GAT CCA CAA GCG GCG GAG TTC AAG TCA
 A   D   G   S   G   D   V   D   P   Q   A   A   E   F   K   S>
```

Figure 7a

```
      680         690         700         710         720
      *   *   *   *   *   *   *   *   *   *   *   *   *   *
ATG GTG GTG GAA ATG ATG GTA GTC GCC GGT GTT TTT AAC ATT GGT GAT
 M   V   V   E   M   M   V   V   A   G   V   F   N   I   G   D>

          730         740         750         760
      *   *   *   *   *   *   *   *   *   *   *   *   *
TTT ATT CCG CAA CTT AAT TGG TTA GAT ATT CAA GGT GTA GCC GCT AAA
 F   I   P   Q   L   N   W   L   D   I   Q   G   V   A   A   K>

770         780         790         800         810
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
ATG AAG AAG CTC CAC GCG CGT TTC GAC GCG TTC TTG ACT GAT ATA CTT
 M   K   K   L   H   A   R   F   D   A   F   L   T   D   I   L>

820         830         840         850         860
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
GAA GAG CAT AAG GGT AAA ATT TTT GGA GAA ATG AAA GAT TTG TTG AGT
 E   E   H   K   G   K   I   F   G   E   M   K   D   L   L   S>

    870         880         890         900         910
      *   *   *   *   *   *   *   *   *   *   *   *   *
ACT TTG ATC TCT CTT AAA AAT GAT GAT GCG GAT AAT GAT GGA GGG AAA
 T   L   I   S   L   K   N   D   D   A   D   N   D   G   G   K>

      920         930         940         950         960
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
CTC ACT GAT ACA GAA ATT AAA GCA TTA CTT TTG AAC TTG TTT GTA GCT
 L   T   D   T   E   I   K   A   L   L   L   N   L   F   V   A>

          970         980         990        1000
      *   *   *   *   *   *   *   *   *   *   *   *   *
GGA ACA GAC ACA TCT TCT AGT ACA GTT GAA TGG GCC ATT GCT GAG CTT
 G   T   D   T   S   S   S   T   V   E   W   A   I   A   E   L>

1010        1020        1030        1040        1050
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
ATT CGT AAT TTA AAA ATA CTA GCC CAA GCC CAG CAA GAG ATC GAC AAA
 I   R   N   L   K   I   L   A   Q   A   Q   Q   E   I   D   K>

1060        1070        1080        1090        1100
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
GTC GTT GGA AAA AAC CAG CTA GTT GGC GAA TTG GAC CTA GCC CAA TTG
 V   V   G   K   N   Q   L   V   G   E   L   D   L   A   Q   L>

1110        1120        1130        1140        1150
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
ACA TAC TTG AAA GCT ATA GTC AAG GAA ACC TTT CGG CTT CAT CCA TCA
 T   Y   L   E   A   I   V   K   E   T   F   R   L   H   P   S>

      1160        1170        1180        1190        1200
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
ACC CCT CTT TCA CTT CCT AGA ATT GCA TCT GAG AGT TGT GAG ATC AAT
 T   P   L   S   L   P   R   I   A   S   E   S   C - E   I   N>

          1210        1220        1230        1240
      *   *   *   *   *   *   *   *   *   *   *   *   *
GGC TAT TTC ATT CCA AAA GGC TCA ACG CTT CTC CTT AAT GTT TGG GCC
 G   Y   F   I   P   K   G   S   T   L   L   L   N   V   W   A>

1250        1260        1270        1280        1290
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
ATT GCT CGT GAT CCA AAT GCA TGG GCT GAT CCA TTG GAG TTT AGG CCT
 I   A   R   D   P   N   A   W   A   D   P   L   E   F   R   P>

1300        1310        1320        1330        1340
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
GAA AGG TTT TTG CCA GGA GGT GAG AAG CCC AAA GTT GAT GTC CGT GGG
 E   R   F   L   P   G   G   E   K   P   K   V   D   V   R   G>
```

Figure 7b cont.

```
      1350        1360        1370        1380        1390
         *           *           *           *           *
AAT GAC TTT GAA GTC ATA CCA TTT GGA GCT GGA CGT AGG ATT TGT GCT
 N   D   F   E   V   I   P   F   G   A   G   R   R   I   C   A

      1400        1410        1420        1430        1440
         *           *           *           *           *
GGA ATG AAT TTG GGT ATA CGT ATG GTC CAG TTG ATG ATT GCA ACT TTA
 G   M   N   L   G   I   R   M   V   Q   L   M   I   A   T   L>

      1450        1460        1470        1480
         *           *           *           *           *
ATA CAT GCA TTT AAC TGG GAT TTG GTC AGT GGA CAA TTG CCG GAG ATG
 I   H   A   F   N   W   D   L   V   S   G   Q   L   P   E   M>

1490        1500        1510        1520        1530
   *           *           *           *           *
TTG AAT ATG GAA GAA GCA TAT GGG CTG ACC TTA CAA CGG GCT GAT CCA
 L   N   M   E   E   A   Y   G   L   T   L   Q   R   A   D   P>

1540        1550        1560        1570        1580
   *           *           *           *           *
TTG GTT GTG CAC CCA AGG CCT CGC TTA GAA GCC CAA GCG TAC ATT GGG
 L   V   V   H   P   R   P   R   L   E   A   Q   A   Y   I   G>

      1590        1600        1610        1620        1630
         *           *           *           *           *
TGA gca gca aca gcc cat gga gat aac atg agt gtt aaa tgt atg agt

      1640        1650        1660        1670        1680
         *           *           *           *           *
ctc cat atc ttg ttt agt ttg ttt atg ctt tgg att tag tag ttt tta

      1690        1700        1710        1720
         *           *           *           *           *
tat tga tag atc aat gtt tgc att gtc agt aag aat atc cgt tgc ttg

1730        1740        1750        1760        1770
   *           *           *           *           *
ttt cat taa ctc cag gtg gac aat aaa aga agt aat ttg tat gaa aaa

1780
   *           *
aaa aaa aaa aaa
```

Figure 7c

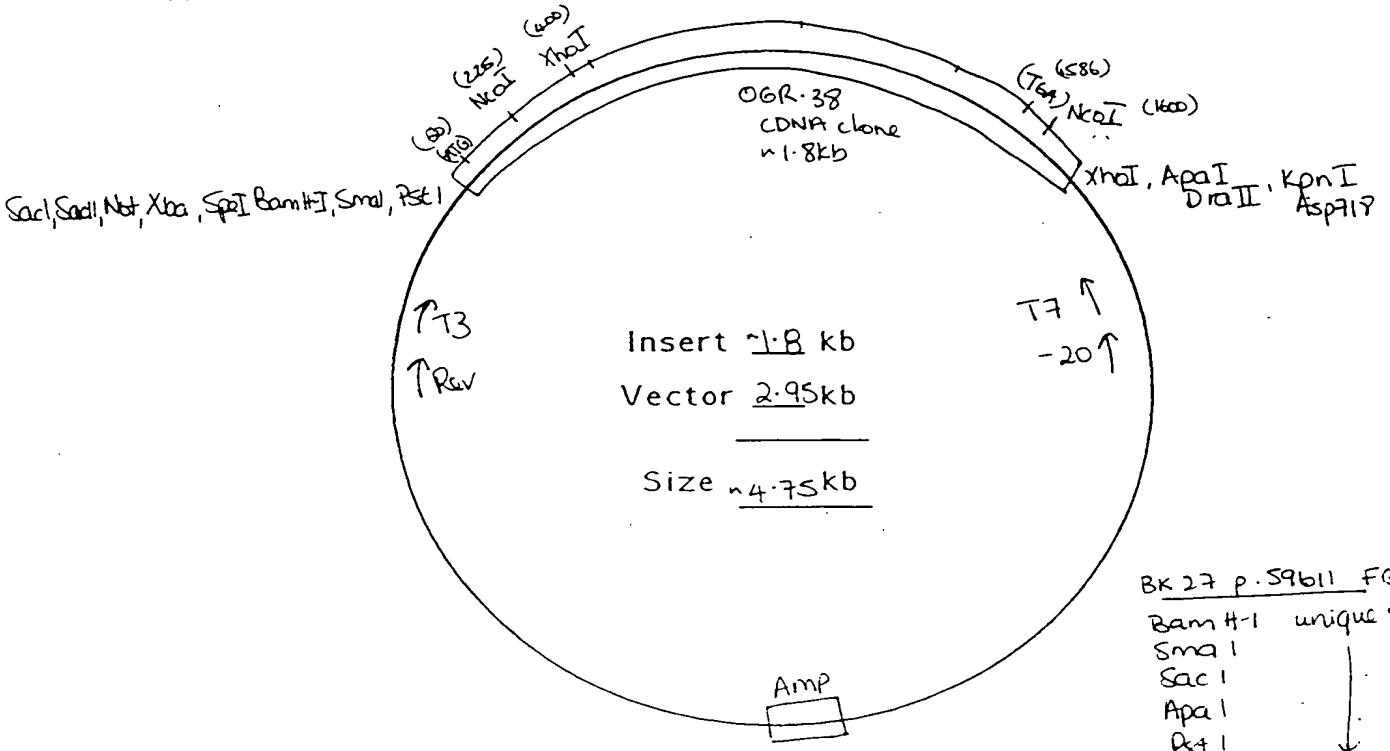
PLASMID pCGP1805

Constructed by Gina Barri-Rewell (OGR cDNA library) Date 17/7/95

Purpose Petunia Ht1 cDNA clone

Phenotype Amp^R

MAP



BK 27 p. 59611 FB

Bam H-1	unique sites
Sma I	
Sac I	
Apa I	
Pst I	
KpnI	
Sal I	no site
XhoI	3kb + 1.45kb fragments

Replicon/Site pBluescript SK II / λZAP II

Insert source cDNA from OGR petals stages 1→3.

Lab Book pp GBR BK 17 and FB BK 26

Comments cDNA library screened with a mixture of P450 fragments (C4H, Hf1, Hf2 and 651)

Linked to Ht1 locus of P. hybrida

Figure 8

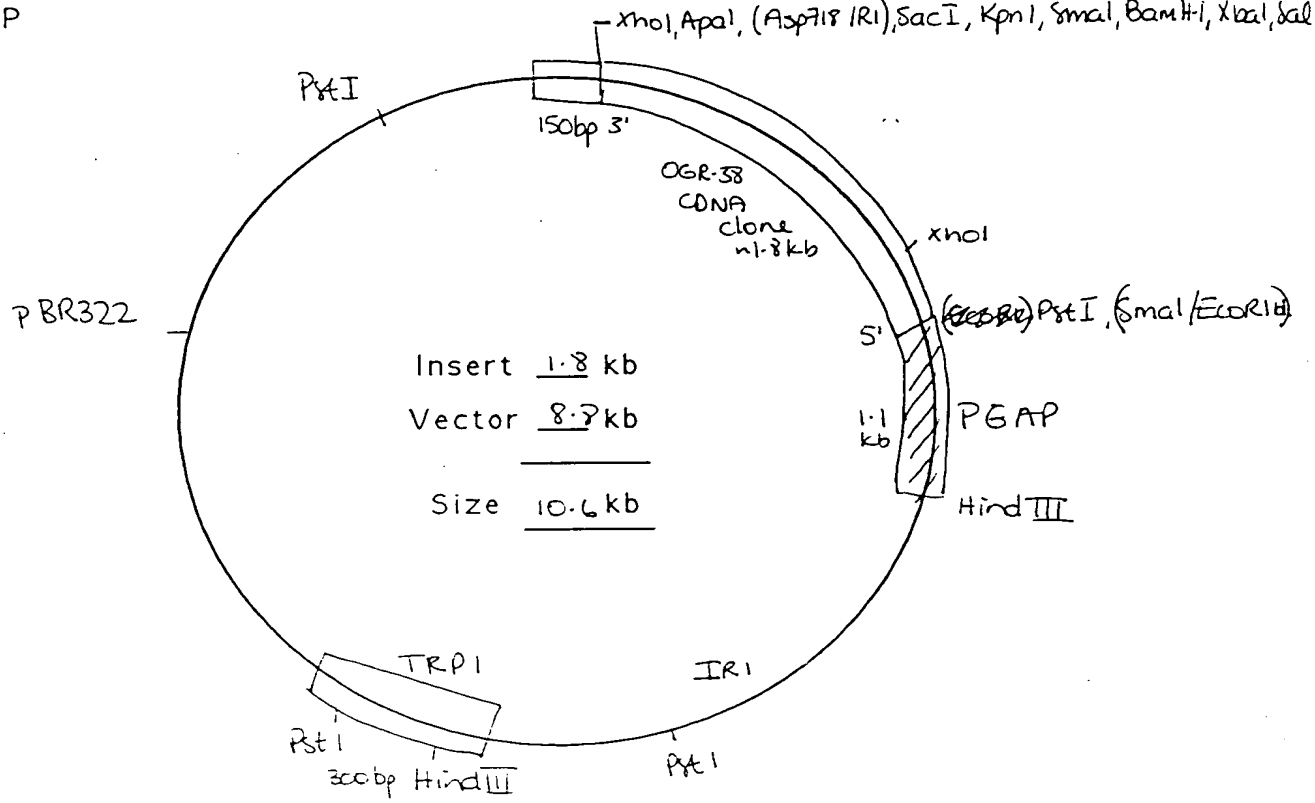
PLASMID pCGP1646

Constructed by Filippa Brugliera Date 3/8/95

Purpose To express OGR-38 cDNA clone in yeast

Phenotype Amp^R in E. coli (TRP1 in yeast)

MAP



Replicon/Site pYE22m EcoRI (blunted)

Insert source 1.8kb Asp718(blunted)/SmaI fragment (OGR-38) from pCGP1805

Lab Book pp BK 27 FB

Comments OGR-38 has F3'H activity in yeast.

Figure 9

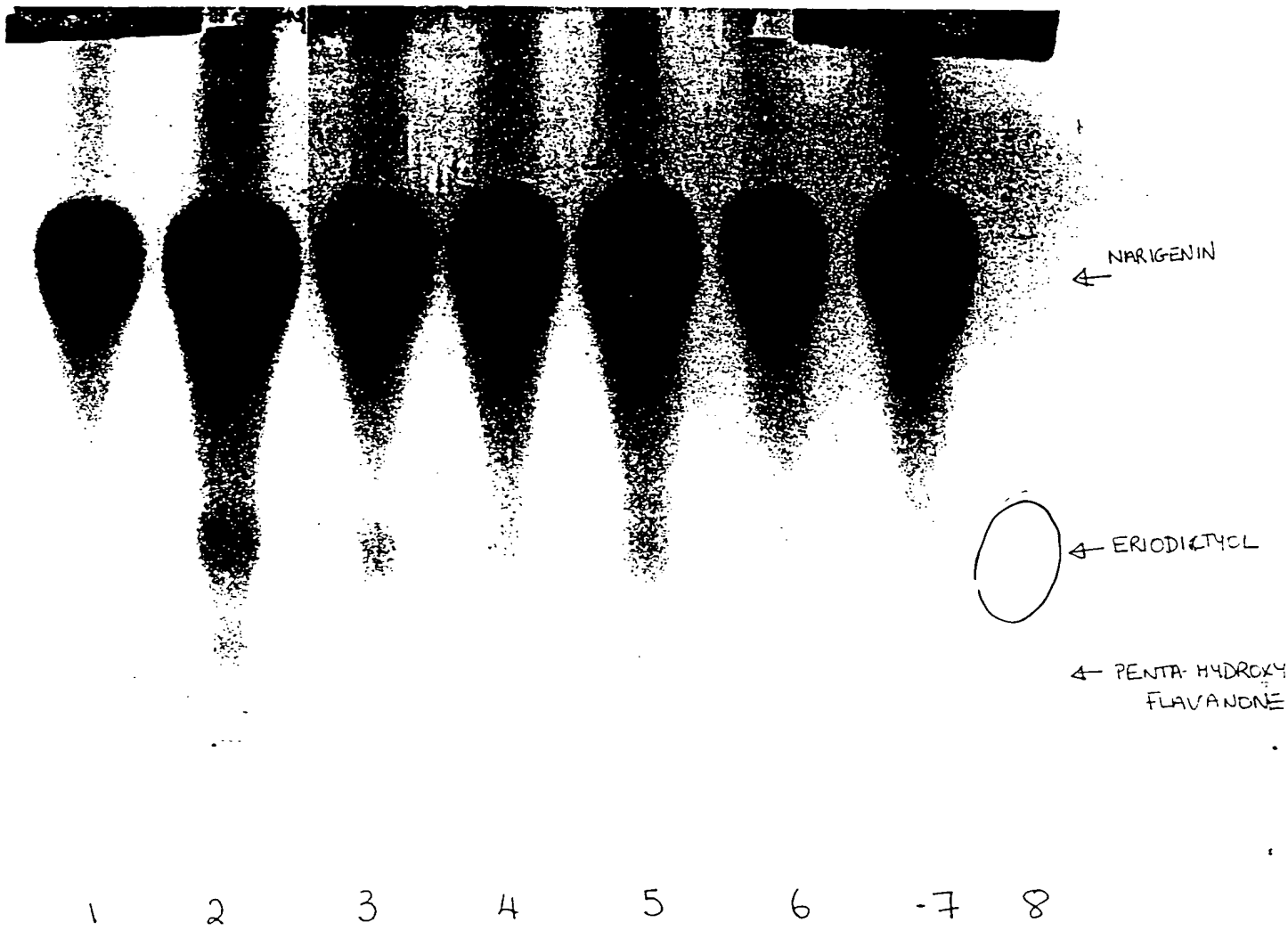


Figure 10

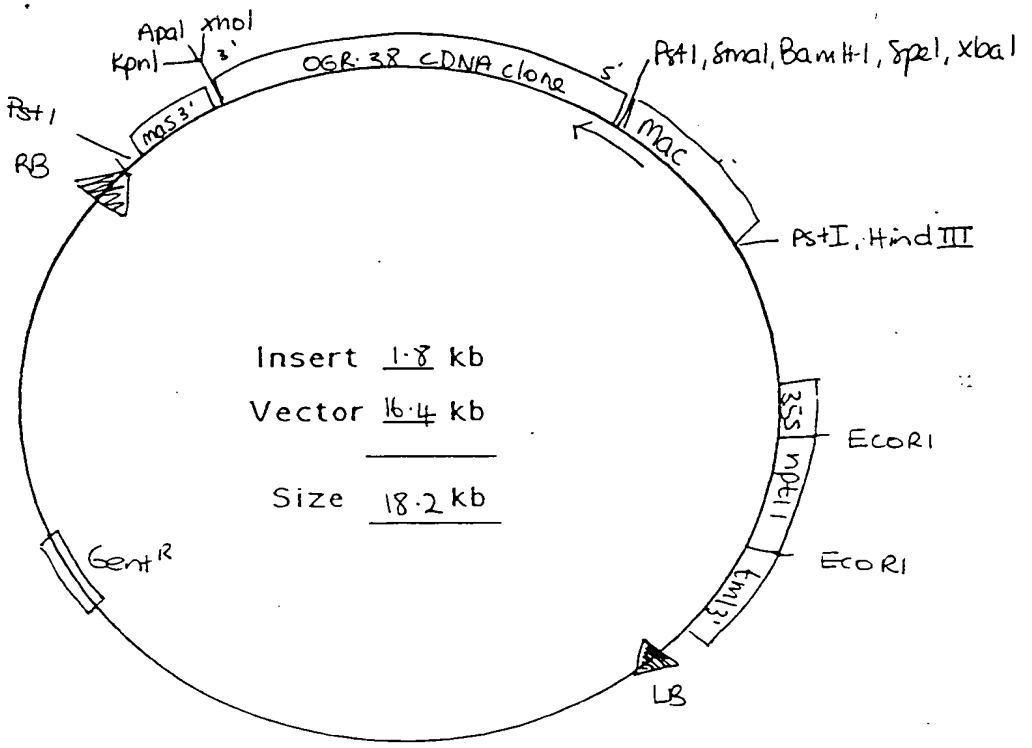
PLASMID pCGP1867

Constructed by G. Barri-Rewell, F. Brugliera, C. Hyland Date 17/8/95

Purpose Expression of Ht1 cDNA clone (OGR·38) in plants

Phenotype Gent^R

MAP



Replicon/Site pCGP293 XbaI | KpnI

Insert source 1·8kb XbaI/KpnI fragment (OGR·38) from pCGP1805

Lab Book pp CH BK 18 p 59145

Comments OGR·38 under the control of the Mac promoter complemented the ht1 mutation in SKr4 x SW63 (Dec. 1995)

Figure 11

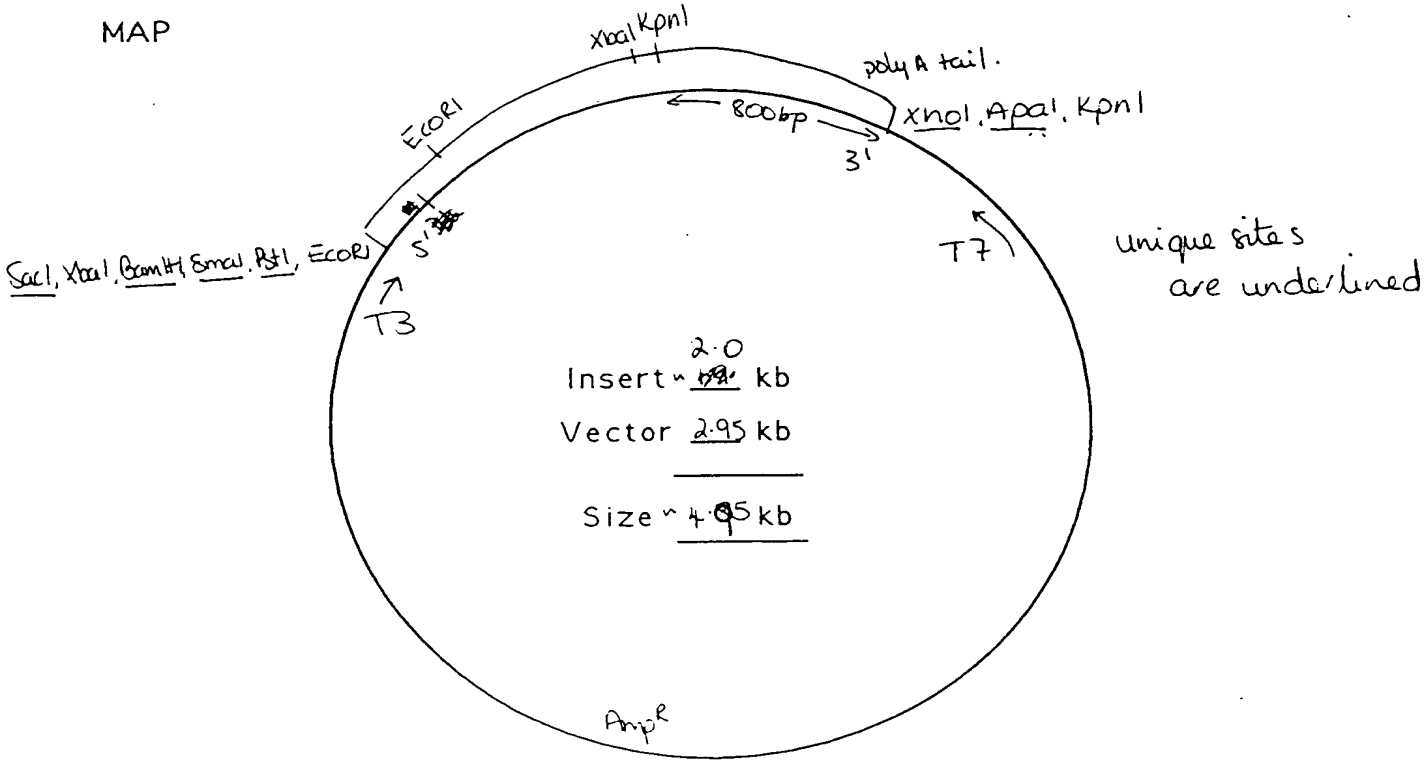
PLASMID pCGP1807

Constructed by Gina Barri-Rewell Date 29/11/95

Purpose F3'H cDNA clone (KC-1) from carnation

Phenotype Amp^R

MAP



Replicon/Site pBluescript II SK- /λZAPII

Insert source cDNA prepared from Kortina Chanel petals (stages 1-3)

Lab Book pp GBR

Comments

Figure 12

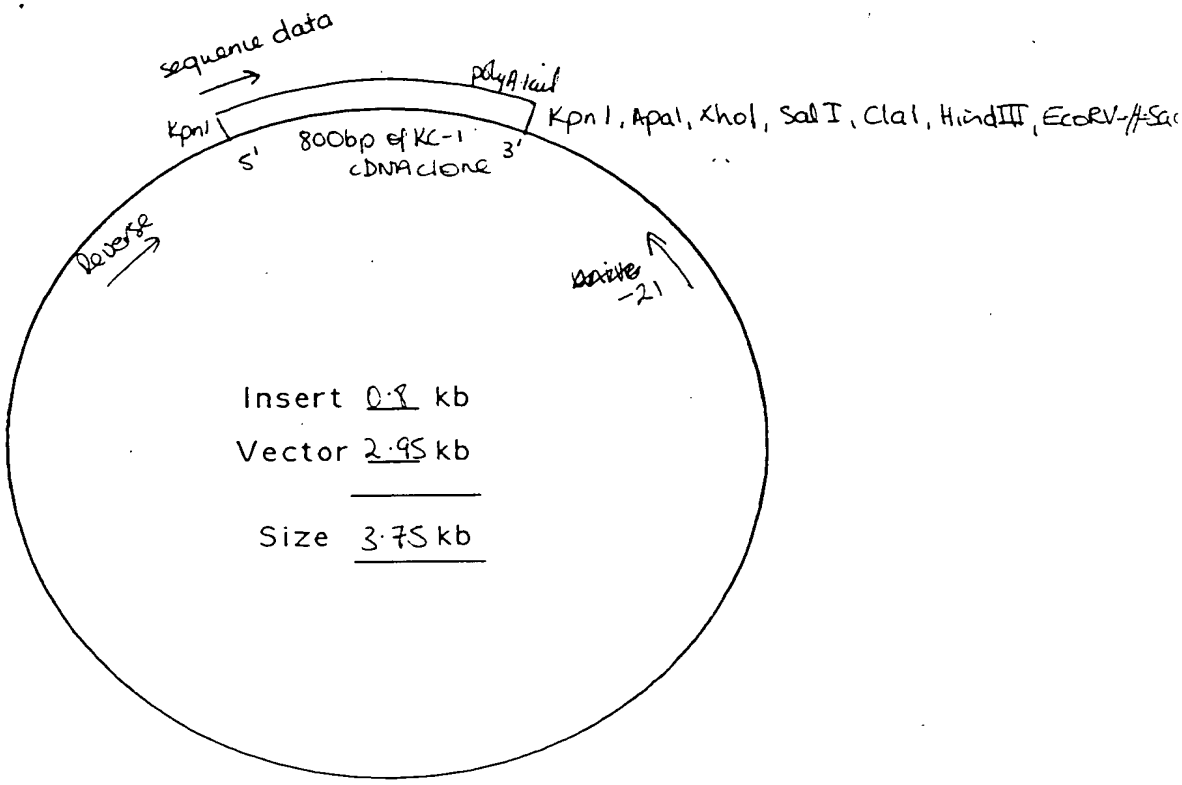
PLASMID pCGP1808

Constructed by Gina Barri-Rowell Date 11-12-95

Purpose Sub-clone of KC-1 fragment to provide sequence data towards the 3' end.

Phenotype Amp^R

MAP



Replicon/Site pBluescript KS II

Insert source 800bp Kpn I fragment from pCGP 1807

Lab Book pp GBR BK 18 p 060450.

Comments

Figure 13

KC-1 (KpnI site) 3'end of cDNA clone Sequence Range: 1 to 458

```
          10         20         30         40
          *          *          *          *          *          *
CAC ATT CCC AAA AAC ACC ACT TTA TTG GTA AAT GTG TGG GCC ATC GCA
 H   I   P   K   N   T   T   L   L   V   N   V   W   A   I   A>

50         60         70         80         90
*          *          *          *          *          *          *
CGC GAC CCT GAG GTT TGG GCC GAC CCG TTA GAG TTT AAA CCC GAA AGA
 R   D   P   E   V   W   A   D   P   L   E   F   K   P   E   R>

  100        110        120        130        140
  *          *          *          *          *          *          *
TTT TTG CCG GGC GGC GAA AAG CCC AAT GTG GAT GTG AAA GGA AAC GAT
 F   L   P   G   G   E   K   P   N   V   D   V   K   G   N   D>

    150        160        170        180        190
*   *          *          *          *          *          *          *
TTT GAG CTG ATT CCG TTC GGG GCG GGC CGA CGG ATT TGT GCT GGG CTG
 F   E   L   I   P   F   G   A   G   R   R   I   C   A   G   L>

        200        210        220        230        240
  *          *          *          *          *          *          *
AGT TTG GGC CTG CGT ATG GTC CAG TTG ATG ACA GCC ACT TTG GCC CAT
 S   L   G   L   R   M   V   Q   L   M   T   A   T   L   A   H>

          250        260        270        280
          *          *          *          *          *          *
ACT TAT GAT TGG GCC TTA GCT GAT GGG CTT ATG CCC GAA AAG CTT AAC
 T   Y   D   W   A   L   A   D   G   L   M   P   E   K   L   N>

290        300        310        320        330
*          *          *          *          *          *          *
ATG GAT GAG GCT TAT GGG CTT ACC TTA CAG CGT AAG GTG CCA CTT AAT
 M   D   E   A   Y   G   L   T   L   Q   R   K   V   P   L   N>

   340        350        360        370        380
   *          *          *          *          *          *          *
GAT CTA CCT GAC CCT GTC GGC TTC TCG GCC CGT GTT taa taa ttc cgg
 G   P   P   D   P   V   G   F   S   A   R   V

    390        400        410        420        430
*   *          *          *          *          *          *          *
ggt ttt taa aag cgg gtt act ttt gtt tat gta tta ttc cgt act agt

        440        450
  *          *          *          *          *
ttg aaa ata atg gta tta gag aaa tg
```

Figure 14

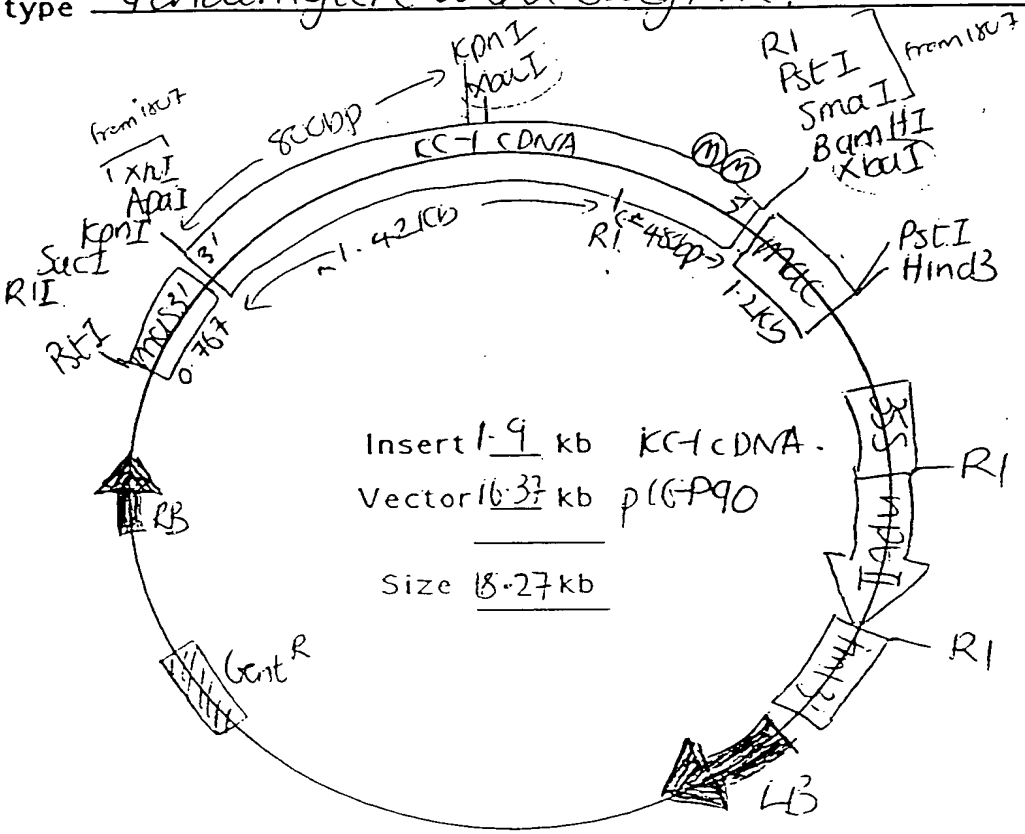
PLASMID pCGP 1800 1810

Constructed by Greg Barrett Revell Date 12-12-95

Purpose Binary vector for KC-1 expression - stable + transient.

Phenotype Gentamycin med 3ug/ml.

MAP



Replicon/Site pCGP293

Insert source BamHI/ApaI fragment of pCGP1807 cloned into BamHI/ApaI sites of pCGP90.

Lab Book pp Book 18 page 060 i-f 1

Comments Checked clones by digesting with ~~EcoRI~~ ApaI / BamHI See Book 18 page 060496

Figure 15

1810 DH5α S/S 24/11/96
1810 LBA S/S 24/11/96

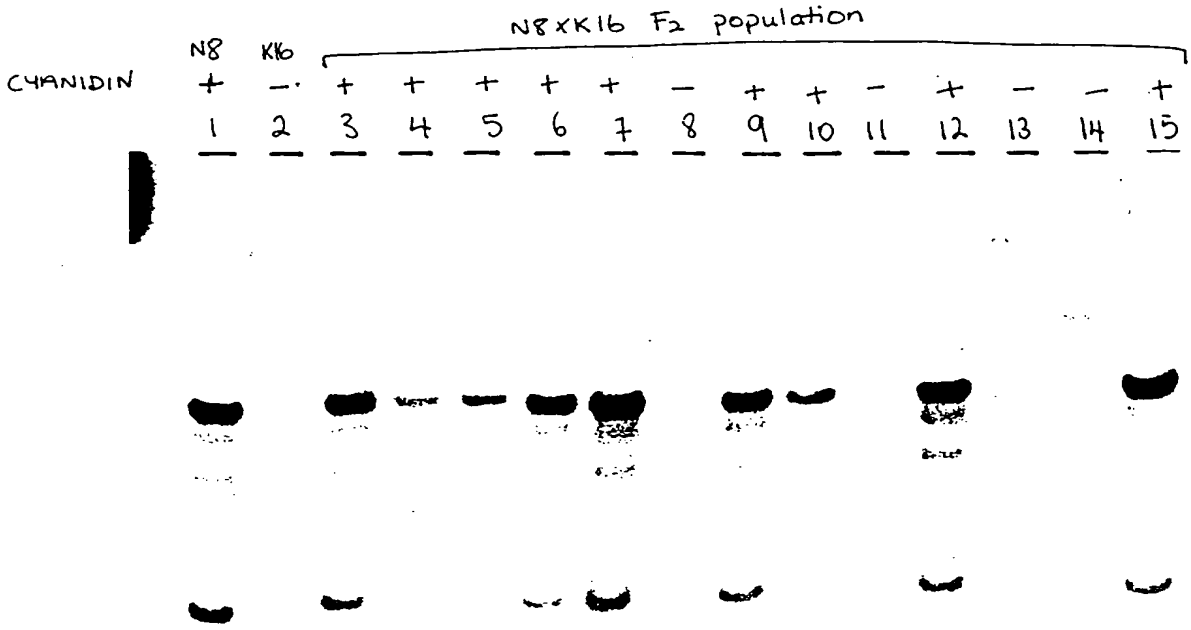

N8xK16 F2 population
N8
K16
CYANIDIN
+ − + + + + + − + + − + − − +
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Figure 16

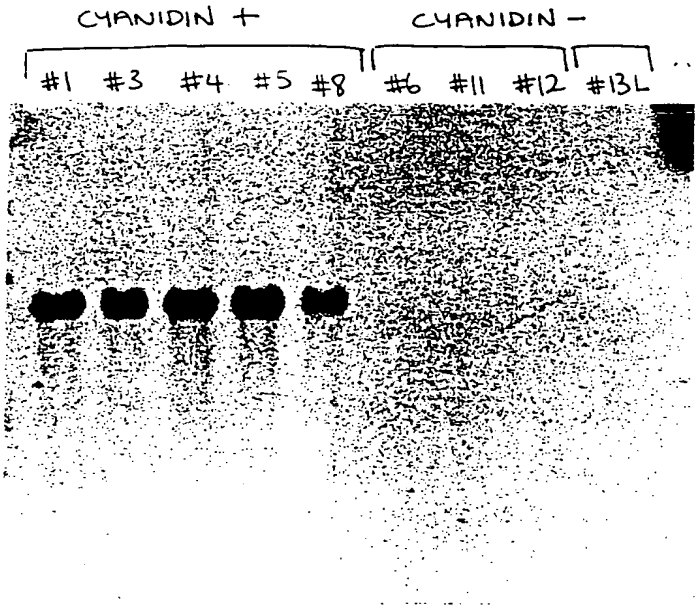
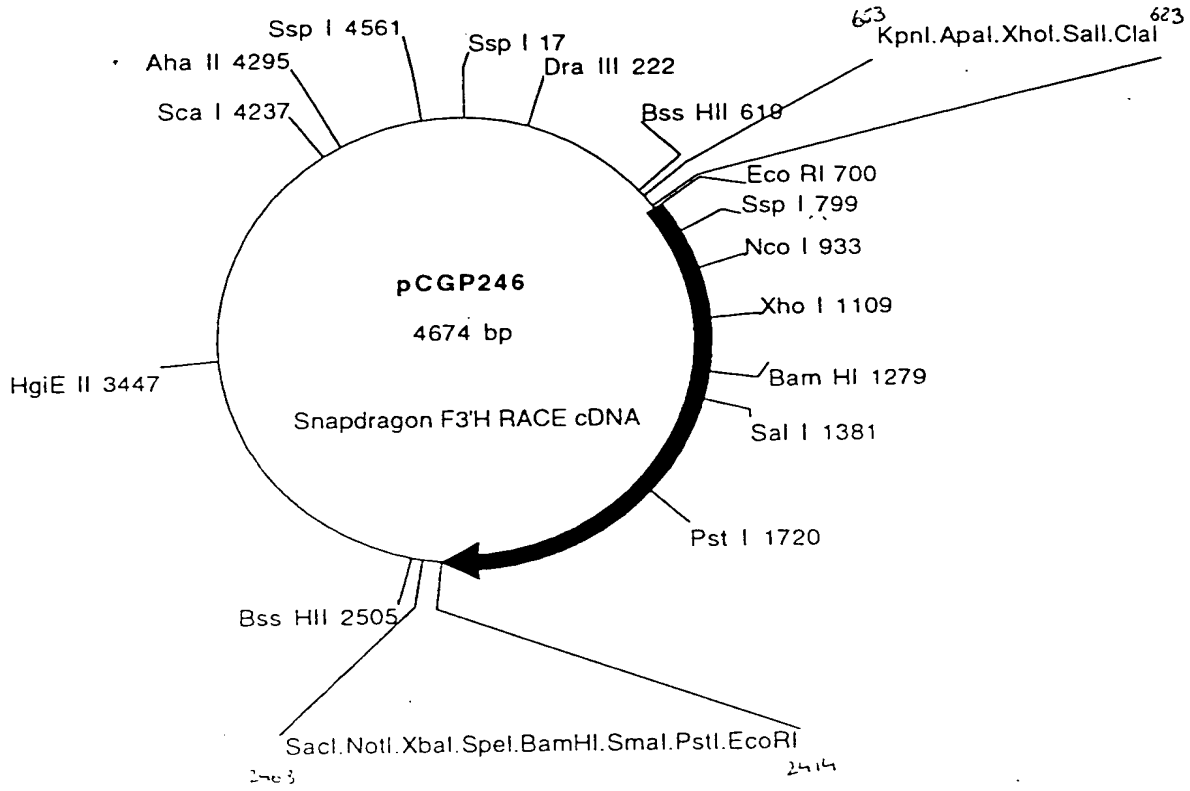


Figure 17



Plasmid name: pCGP246
Plasmid size: 4674 bp
Constructed by: Michael Michael
Construction date: 26/9/95
Comments/References: RACE cDNA of Snapdragon F3'H cloned into T-tailed EcoRV linearised pBluescript II SK+.

Figure 18

Antirrhinum majus PS H cDNA clone

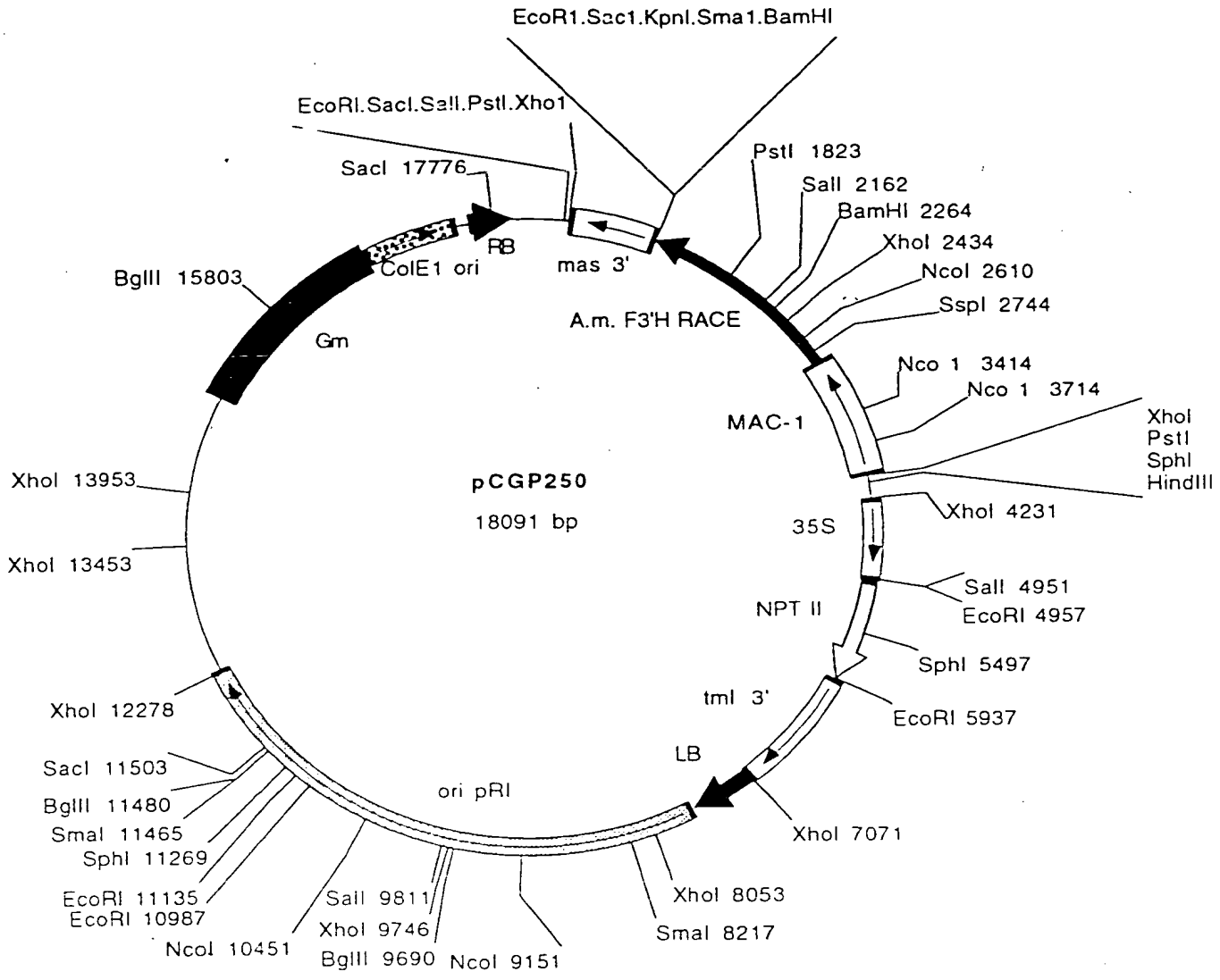
Sequence Range: 1 to 1711

```
          10          20          30          40          50          60
cga att ccc ccc ccc cca cac cat tca atg cct aag tcc tcc att tgc cgg cct aat aac taa aag

   70          80          90         100         110         120         130
ccc act ctt tcc gac cat cta tac ATG CAA CAC CAA TAT TAT TCT TTA ATT ACG ATG GAT GAT ATT
                               M   Q   H   Q   Y   Y   S   L   I   T   M   D   D   I>

       140         150         160         170         180         190
AGC ATA ACC AGC TTA TTG GTG CCA TGT ACT TTT ATA TTA GGG TTC TTG CTT CTA TAT TCC TTC CTC
 S   I   T   S   L   L   V   P   C   T   F   I   L   G   F   L   L   L   Y   S   F   L>

200         210         220         230         240         250         260
AAC AAA AAA GTA AAG CCA CTG CCA CCT GGA CCG AAG CCA TGG CCC ATC GTC GGA AAT CTG CCA CAT
 N   K   K   V   K   P   L   P   P   G   P   K   P   W   P   I   V   G   N   L   P   H>

    270         280         290         300         310         320         330
CTT GGG CCG AAG CCC CAC CAG TCG ATG GCG GCG CTG GCA CGG GTG CAC GGC CCA TTA ATT CAT CTG
 L   G   P   K   P   H   Q   S   M   A   A   L   A   R   V   H   G   P   L   I   H   L>

          340         350         360         370         380         390
AAG ATG GGC TTT GTG CAT GTG GTT GTG GCC TCC TCA GCA TCC GTT GCG GAG AAA TTT CTG AAG GTG
 K   M   G   F   V   H   V   V   V   A   S   S   A   S   V   A   E   K   F   L   K   V>

  400         410         420         430         440         450         460
CAT GAC GCA AAC TTC TCG AGC AGG CCT CCC AAT TCG GGT GCA AAA CAC GTG GCC TAC AAC TAT CAG
 H   D   A   N   F   S   S   R   P   P   N   S   G   A   K   H   V   A   Y   N   Y   Q>

       470         480         490         500         510         520
GAC TTG GTC TTT GCT CCT TAT GGC CCA CGC TGG CGG ATG CTC AGG AAA ATC TGT GCA CTC CAC CTC
 D   L   V   F   A   P   Y   G   P   R   W   R   M   L   R   K   I   C   A   L   H   L>

530         540         550         560         570         580         590
TTC TCC GCC AAA GCC TTG AAC GAC TTC ACA CAC GTC AGA CAG GAT GAG GTG GGG ATC CTC ACT CGC
 F   S   A   K   A   L   N   D   F   T   H   V   R   Q   D   E   V   G   I   L   T   R>

    600         610         620         630         640         650         660
GTT CTA GCA GAT GCA GGA GAA ACG CCG TTG AAA TTA GGG CAG ATG ATG AAC ACA TGC GCC ACC AAT
 V   L   A   D   A   G   E   T   P   L   K   L   G   Q   M   M   N   T   C   A   T   N>

          670         680         690         700         710         720
GCA ATA GCG CGT GTT ATG TTG GGT CGA CGC GTG GTT GGA CAC GCA GAC TCA AAG GCG GAG GAG TTT
 A   I   A   R   V   M   L   G   R   R   V   V   G   H   A   D   S   K   A   E   E   F>

  730         740         750         760         770         780         790
AAG GCA ATG GTA GTG GAG TTG ATG GTA TTA GCT GGT GTG TTC AAC TTA GGT GAT TTT ATC CCA CCT
 K   A   M   V   V   E   L   M   V   L   A   G   V   F   N   L   G   D   F   I   P   P>

       800         810         820         830         840         850
CTT GAA AAA TTG GAT CTT CAA GGT GTC ATT GCT AAG ATG AAG AAG CTT CAC TTG CGT TTC GAC TCG
 L   E   K   L   D   L   Q   G   V   I   A   K   M   K   K   L   H   L   R   F   D   S>

860         870         880         890         900         910         920
TTC TTG AGT AAG ATC CTT GGA GAC CAC AAG ATC AAC AGC TCA GAT GAA ACC AAA GGC CAT TCG GAT
 F   L   S   K   I   L   G   D   H   K   I   N   S   S   D   E   T   K   G   H   S   D>

    930         940         950         960         970         980         990
TTG TTG AAC ATG TTA ATT TCT TTG AAG GAC GCT GAT GAT GCC GAA GGA GGG AGG CTC ACC GAC GTA
 L   L   N   M   L   I   S   L   K   D   A   D   D   A   E   G   G   R   L   T   D   V>

         1000        1010        1020        1030        1040        1050
GAA ATT AAA GCG TTG CTC TTG AAC TTG TTT GCT GCA GGA ACT GAC ACA ACA TCA AGC ACT GTG GAA
 E   I   K   A   L   L   L   N   L   F   A   A   G   T   D   T   T   S   S   T   V   E>

  1060        1070        1080        1090        1100        1110        1120
TGG TGC ATA GCT GAG TTA GTA CGA CAT CCT GAA ATC CTT GCC CAA GTC CAA AAA GAA CTC GAC TCT
 W   C   I   A   E   L   V   R   H   P   E   I   L   A   Q   V   Q   K   E   L   D   S>

       1130        1140        1150        1160        1170        1180
GTT GTT GGT AAG AAT CGG GTG GTG AAG GAG GCT GAT CTG GCC GGA TTA CCA TTC CTC CAA GCG GTC
 V   V   G   K   N   R   V   V   K   E   A   D   L   A   G   L   P   F   L   Q   A   V>
```

Figure 19a

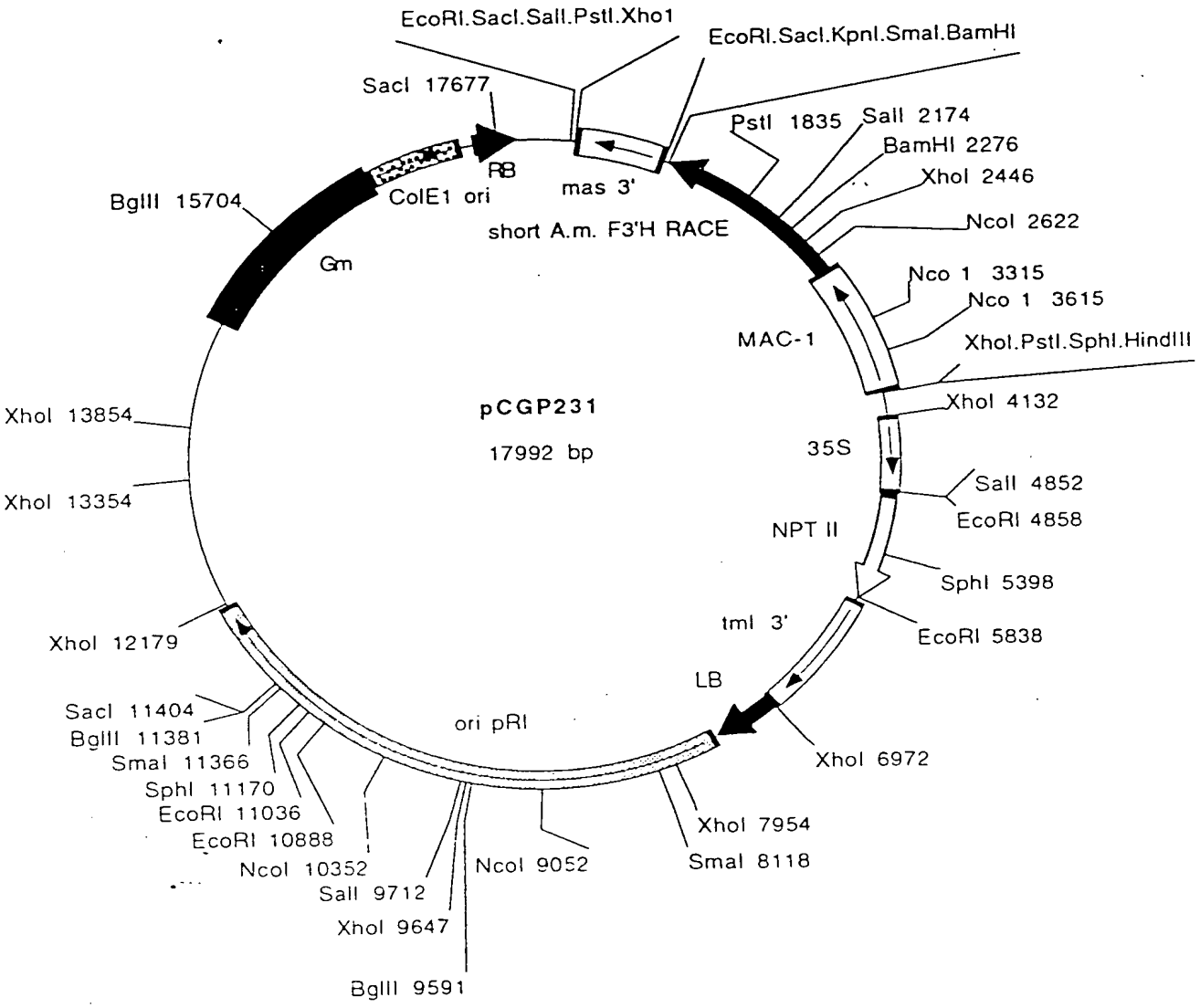
```
1190        1200        1210        1220        1230        1240        1250
 .    .    .    .    .    .    .    .    .    .    .    .    .    .
GTC AAG GAA AAT TTC CGA CTC CAT CCC TCC ACC CCG CTC TCC CTA CCG AGG ATC GCA CAT GAG AGT
 V   K   E   N   F   R   L   H   P   S   T   P   L   S   L   P   R   I   A   H   E   S>

   1260        1270        1280        1290        1300        1310        1320
 .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .
TGT GAA GTG AAT GGA TAC TTG ATT CCA AAG GGT TCG ACA CTT CTT GTC AAT GTT TGG GCA ATT GCT
 C   E   V   N   G   Y   L   I   P   K   G   S   T   L   L   V   N   V   W   A   I   A>

        1330        1340        1350        1360        1370        1380
 .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .
CGC GAT CCA AAT GTG TGG GAT GAA CCA CTA GAG TTC CGG CCT GAA CGA TTC TTG AAG GGC GGG GAA
 R   D   P   N   V   W   D   E   P   L   E   F   R   P   E   R   F   L   K   G   G   E>

 1390        1400        1410        1420        1430        1440        1450
 .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .
AAG CCT AAT GTC GAT GTT AGA GGG AAT GAT TTC GAA TTG ATA CCG TTC GGA GCG GGC CGA AGA ATT
 K   P   N   V   D   V   R   G   N   D   F   E   L   I   P   F   G   A   G   R   R   I>

      1460        1470        1480        1490        1500        1510
 .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .
TGT GCA GGA ATG AGC TTA GGA ATA CGT ATG GTC CAG TTG TTG ACA GCA ACT TTG AAC CAT GCG TTT
 C   A   G   M   S   L   G   I   R   M   V   Q   L   L   T   A   T   L   N   H   A   F>

1520        1530        1540        1550        1560        1570        1580
 .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .
GAC TTT GAT TTG GCG GAT GGA CAG TTG CCT GAA AGC TTA AAC ATG GAG GAA GCT TAT GGG CTG ACC
 D   F   D   L   A   D   G   Q   L   P   E   S   L   N   M   E   E   A   Y   G   L   T>

   1590        1600        1610        1620        1630        1640        1650
 .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .
TTG CAA CGA GCT GAC CCT TTG GTA GTG CAC CCG AAG CCT AGG TAG gca cct cat gtt tat caa act
 L   Q   R   A   D   P   L   V   V   H   P   K   P   R   *

        1660        1670        1680        1690        1700        1710
 .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .
tag gac tca tgt tta gag aac ctc ttg ttg ttt tat cag att gaa gtg tga tgt cca aga c
```

Figure 19b



Plasmid name: pCGP250
Plasmid size: 18091 bp
Constructed by: Michael Michael
Construction date: 23/11/95
Comm nts/R f rences: Blunted Eco RI insert from pCGP246 (Snapdragon F3'H RACE cDNA) cloned into blunted Xba I linearised pCGP293; npt = Tn5 Kan , fragment sizes are only rough

Figure 20



Plasmid name: pCGP231
Plasmid size: 17992 bp
Constructed by: Michael Michael
Construction date: 4/12/95
Comm nts/Ref rences: SmaI/SspI insert of pCGP246 [short (1 ATG) Snapdragon F3'H RACE cDNA] cloned into blunted XbaI linearised pCGP293; npt = Tn5 Kan , fragment sizes are only rough

Figure 21

Arab

Sequence Range: 1 to 971

```
                10                  20                  30                  40
GAT ATG CTT AGC ACT TTA ATC TCC CTT AAA GGA ACT GAT CTT GAC
 D   M   L   S   T   L   I   S   L   K   G   T   D   L   D>

    50                  60                  70                  80                  90
GGT GAC GGA GGA AGC TTA ACG GAT ACT GAG ATT AAA GCC TTG CTA
 G   D   G   G   S   L   T   D   T   E   I   K   A   L   L>

            100                 110                 120                 130
                *
TTG AAC ATG TTC ACA GCT GGA ACT GAC ACG TCA GCA AGT ACG GTG
 L   N   M   F   T   A   G   T   D   T   S   A   S   T   V>

   140                 150                 160                 170                 180
GAC TGG GCT ATA GCT GAA CTT ATC CGT CAC CCG GAT ATA ATG GTT
 D   W   A   I   A   E   L   I   R   H   P   D   I   M   V>

            190                 200                 210                 220
                                *
AAA GCC CAA GAA GAA CTT GAT ATT GTT GTG GGC CGT GAC AGG CCT
 K   A   Q   E   E   L   D   I   V   V   G   R   D   R   P>

   230                 240                 250                 260                 270
GTT AAT GAA TCA GAC ATC GCT CAG CTT CCT TAC CTT CAG GCG GTT
 V   N   E   S   D   I   A   Q   L   P   Y   L   Q   A   V>

            280                 290                 300                 310
                                                    *
ATC AAA GAG AAT TTC AGG CTT CAT CCA CCA ACA CCA CTC TCG TTA
 I   K   E   N   F   R   L   H   P   P   T   P   L   S   L>

   320                 330                 340                 350                 360
CCA CAC ATC GCG TCA GAG AGC TGT GAG ATC AAC GGC TAC CAT ATC
 P   H   I   A   S   E   S   C   E   I   N   G   Y   H   I>

            370                 380                 390                 400
                                                                        *
CCG AAA GGA TCG ACT CTA TTT GAC GGA CAT ATG GGC CTA GGC CGT
 P   K   G   S   T   L   F   D   G   H   M   G   L   G   R>

   410                 420                 430                 440                 450
GAC CCG GAT CAA TGG TCC GAC CCG TTA GCA TTT AAA CCC GAG AGA
 D   P   D   Q   W   S   D   P   L   A   F   K   P   E   R>

            460                 470                 480                 490
TTC TTA CCC GGT GGT GAA AAA TCC GGC GTT GAT GTG AAA GGA AGC
 F   L   P   G   G   E   K   S   G   V   D   V   K   G   S>

   500                 510                 520                 530                 540
    *
GAT TTC GAG CTA ATA CCG TTC GGG GCT GGG AGG CCA ATC TGT GCA
 D   F   E   L   I   P   F   G   A   G   R   P   I   C   A>

            550                 560                 570                 580
GGT TTA AGT TTA GGG CTA CGT ACA GAT TTA AGT TGC CTT CAC GCC
 G   L   S   L   G   L   R   T   D   L   S   C   L   H   A>

   590                 600                 610                 620                 630
                       *
AAC GTT GCT CAC AAG CAT TTG ATT GGG AAC TTC AGC TGG AGA AGT
 N   V   A   H   K   H   L   I   G   N   F   S   W   R   S>
```

Figure 22a

```
          640         650         660         670
TAC GCC GGA CAA CCT GAA TAT CGC AGG AAA AGT TTA CTG GGC TTT
 Y   A   G   Q   P   E   Y   R   R   K   S   L   L   G   F>

   680         690         700         710         720
                           *
ACA CTG CAA AGA GCG GTT CCT TCG GTG GTA CAC CCT AAG CCA AGG
 T   L   Q   R   A   V   P   S   V   V   H   P   K   P   R>

          730         740         750         760
TTG GCC CCG AAC GTT TAT GGA CCC CGG GTC GGC TTA AAA TTT AAC
 L   A   P   N   V   Y   G   P   R   V   G   L   K   F   N>

   770         780         790         800         810
                                       *
TTT GCT TCT TGG ACA AGG TAT ATG GCT TGC ACG AAA CTA ACG TTT
 F   A   S   W   T   R   Y   M   A   C   T   K   L   T   F>

          820         830         840         850
TAA CAC ACC GTA GTT TGA TCC GGA GTT AGC TTT ATG TAA GAA CGT
 *

   860         870         880         890         900
                                                     *
GTA ACG CCA AAT CAA GCC ATT ATC AAC TAC CGT GAG CTG TTT GTA

          910         920         930         940
CCC TAT CTA TAA ATC TTG AAG AGG AAC ATT TCA GAA CTC TTG ACT

   950         960         970
ATG TTT CAG GAA CAA AAA AAA AAA AA
```

Figure 22b

B4 5' end

```
          10          20          30          40
      *   *   *   *   *   *   *   *   *   *
A GCG CAT GCC TTG GCA AAT TCA GGG TCA AAG GTA GTG AAC CTG GCG CAA
  A   H   A   L   A   N   S   G   S   K   V   V   N   L   A   Q>

50          60          70          80          90
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
CTG CTG AAC CTG TGC ACG GTC AAT GCT CTA GGA AGG GTG ATG GTA GGG
 L   L   N   L   C   T   V   N   A   L   G   R   V   M   V   G>

100         110         120         130         140
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
CGG AGG GTT TTC GGC GAC GGC AGC GGA GGC GAC GAT CCG AAG GCG GAC
 R   R   V   F   G   D   G   S   G   G   D   D   P   K   A   D>

    150         160         170         180         190
    *   *   *   *   *   *   *   *   *   *   *   *   *   *
GAG TTC AAA TCG ATG GTG GTG GAG ATG ATG GTG TTG GCA GGA GTG TTC
 E   F   K   S   M   V   V   E   M   M   V   L   A   G   V   F>

        200         210         220         230         240
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
AAC ATA GGT GAC TTC ATC CCC TCT CTC GAA TGG CTT GAC TTG CAA GGC
 N   I   G   D   F   I   P   S   L   E   W   L   D   L   Q   G>

        250         260         270         280
    *   *   *   *   *   *   *   *   *   *   *   *
GTG GCG TCC AAG ATG AAG AAG CTC CAC AAG AGA TTC GAC GAC TTC TTG
 V   A   S   K   M   K   K   L   H   K   R   F   D   D   F   L>

290         300         310         320         330
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
ACA GCC ATT GTC GAG GAC CAC AAG AAG GGC TCC GGC ACG GCG GGG CAC
 T   A   I   V   E   D   H   K   K   G   S   G   T   A   G   H>

340         350         360         370         380
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
GTC GAC ATG TTG ACC ACT CTG CTC TCG CTC AAG GAA GAC GCC GAC GGC
 V   D   M   L   T   T   L   L   S   L   K   E   D   A   D   G>
```

Figure 23 a

R4 3' end

```
          10          20          30          40
          *   *   *   *   *   *   *   *   *   *
T GCC GGG CGA AGA ATA TGT GCC GGG ATG AGC TTG GGC CTC CGT ATG GTC
  A   G   R   R   I   C   A   G   M   S   L   G   L   R   M   V>

50          60          70          80          90
*   *   *   *   *   *   *   *   *   *   *   *   *   *   *
CAT TTA ATG ACT GCA ACA TTG GTC CAC GCA TTT AAT TGG GCC TTG GCT
 H   L   M   T   A   T   L   V   H   A   F   N   W   A   L   A>

100         110         120         130         140
  *   *   *   *   *   *   *   *   *   *   *   *   *   *
GAT GGG CTG ACC GCT GAG AAG TTA AAC ATG GAT GAA GCA TAT GGG CTC
 D   G   L   T   A   E   K   L   N   M   D   E   A   Y   G   L>

   150         160         170         180         190
    *   *   *   *   *   *   *   *   *   *   *   *   *
ACT CTA CAA CGA GCT GCA CCG TTA ATG GTG CAC CCG CGC ACC AGG CTG
 T   L   Q   R   A   A   P   L   M   V   H   P   R   T   R   L>

      200         210         220         230         240
  *   *   *   *   *   *   *   *   *   *   *   *   *   *
GCC CCA CAG GCA TAT AAA ACT TCA TCA TCT TAA tta gag agc tat gtt
 a   p   q   a   y   k   t   s   s   s   *

        250         260         270         280
    *   *   *   *   *   *   *   *   *   *   *   *
ctg ggt gtg ccc ggt ttg atg tct cca tgt ttt cta ttt agg ttt aaa

300         310         320         330
*   *   *   *   *   *   *   *   *   *   *   *
tct gta aga taa ggt gat tct atg ctg aat cac aaa agt tgc tat cta

340         350         360         370         380
*   *   *   *   *   *   *   *   *   *   *   *   *
aat tct atg tcc aat gaa acc gtt ctt ctt ccc ttc tta taa ttt atg

        400
    *   *   *
aat act tat gaa aaa
```

Figure 23b